AN ELECTRONIC SYSTEM FOR WEAR-DEBRIS CONDITION MONITORING

Thesis submitted by

IAN MacPHERSON FLANAGAN

for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Science

University of Edinburgh

March 1987

DECLARATION

I declare that this thesis is my original work except where stated otherwise.

I M Flanagan

CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Dr H W Whittington and Dr J R Jordan, for their support and interest in this work.

The financial assistance of the Science and Engineering Research Council is gratefully acknowledged as is the CASE contribution provided by Tecalemit Electronics Ltd and Gabriel Microwave Systems Ltd. Also, I wish to acknowledge Mr R Bogue and Mr G Smith, formerly of Tecalemit Ltd, whose help during the initial stages of the project was invaluable.

Throughout the course of the work many members of the academic staff of the Department of Electrical Engineering have spent time discussing various aspects of the project with me. I would like to thank them generally and, in particular, Drs Dinnis, Dripps, Dryburgh, Flynn and Grant for detailed discussions on various topics. In addition, I wish to acknowledge the advice and kindness of Professor A E Owen.

The use of Departmental computing facilities has made much of this work feasible and I would like to express my gratitude to Dr I A B Lindsay and Mr K M Farvis for their help in this connection.

I am indebted to Dr D Mills of the Department of Mechanical Engineering for arranging the manufacture and supply of material samples.

I wish to thank the technical staff of the Department of Electrical Engineering and, particularly, the assistance of Mr S King, Mr R Addison and Mr R Stevens for their help with constructing apparatus and circuitry, Mr J Goodall for expertly producing SEM photographs of samples, and Mr W L Hillam for his invaluable help in obtaining components and equipment.

I should also like to thank the Engineering Librarian, Mrs V H Thomson, whose efforts and skill have been much appreciated.

Finally, I should like to thank my friends and colleagues, particularly Joaquim DeLima, Neil Petrie, Wee Kiong Choi and Keith Sylvan, for their continual support over the years.

ABSTRACT

Wear-debris condition monitoring involves detecting and identifying wear-debris particles in the lubricant flow of a machine to obtain an estimate of the machine's 'health'. In this work a wear-debris condition monitoring system using an inductive transducer has been investigated and a prototype system has been developed, constructed and successfully tested. The general field of condition monitoring has been reviewed and the more specific area of lubricant wear-debris monitoring has been examined in detail.

The wear-debris monitoring system comprises an inductive transducer, a marginal oscillator, and AM and FM demodulation circuitry. A wear-debris particle passing through the transducer interacts with the system to generate a pair of output pulses; an estimate of the particle diameter and material is obtained from analysing pulse heights. The theoretical foundations of the present method are described in detail.

The system has been tested using particles of lead, brass, copper, steel and ferrite, in the approximate diameter range 100μm to 400μm, and the experimental results confirm the theoretical predictions. Limitations of the present system have been identified, a major problem being noise which limits the minimum detectable size of particle to $\sim$100μm diameter for ferrous materials and $\sim$200μm diameter for non-ferrous. Also, distortion limits the maximum diameter of particle which can be analysed to $\sim$400μm for ferrous materials and $\sim$500μm for non-ferrous.

A comparison between the performance of the present system and the characteristics of commercially available wear-debris monitoring systems is made with the conclusion that the present system offers significant advantages.

Finally, possible future improvements for the system are identified and a recommended programme for further work is given.

CHAPTER ONE

INTRODUCTION

In this chapter we introduce the field of condition monitoring and include a brief review of the range of techniques available for its implementation.

1.1 CONDITION MONITORING AND ITS VALUE TO INDUSTRY

Condition monitoring involves the measurement of certain parameters associated with a particular item of plant which are indicative of the condition of the plant, and the use of the data obtained to assess the state of health of the plant. Maintenance may then be carried out in reaction to implied machine condition, rather than on an estimate based on machine use or the time interval since the previous maintenance action. In comparison with traditional methods there are substantial benefits to such a philosophy.

Traditionally, maintenance programmes have been of two basic types. Breakdown maintenance, based on running machines to failure, then repairing, is clearly a crude approach which may be economic in a few situations, but is not generally acceptable to industry. A better method, regular preventive maintenance, involves stopping machines after regular intervals of time, although selection of an appropriate interval is difficult. A balance is required between operating with a relatively short interval, where many scheduled machine stoppages will not require to be followed by any maintenance, and operating with a much longer interval where there will be a comparatively high risk of machine failure. Even with a short interval, however, there will still be a finite probability of machine failure.

On the other hand, a condition-monitoring based maintenance programme does not depend on selection of a maintenance interval but is obtained directly from machine condition. This implies that the information obtained through condition monitoring techniques is accurate, and analytic algorithms are valid: if this is so, there is little chance of unexpected failure.

When the machine condition is assessed, the intervals between assessments must normally be short in comparison with the expected time between normal and failure conditions. Using this information, in conjunction with historical records for the machine in question as well as machines of the same type, we may estimate the machine's 'health'. Trends may be observed from data collected, and predictions made on future condition.

Condition monitoring measurement techniques display features associated with both off-line and on-line operation. Off-line techniques involve monitoring some machine parameter by removing a characteristic sample to a laboratory for analysis. This is usually slow, expensive and requires skilled personnel, but may still be economic in some cases. On-line techniques, on the other hand, are in principle immediate, usually inexpensive to operate, and, as such, should offer the closest realisation of an ideal condition monitoring scheme, where machine condition is known continuously.

The principle motivation for introducing a condition monitoring programme into industry is economic, although in some cases safety may also be an important consideration. The financial benefits which are normally identified are due to reduced levels of lost production, more convenient maintenance scheduling and reduced repair costs through faulty components causing damage to previously good components. Costs incurred in setting up a condition monitoring programme include installation as well as capital equipment and maintenance costs for the monitoring system. Installation costs are much reduced if monitoring equipment is manufactured as an integral part of the machine. These costs are normally outweighed by the benefits, especially for the low capital, installation and maintenance cost monitoring systems under development today.

1.2 CONDITION MONITORING TECHNIQUES

In the previous section we discussed condition monitoring without referring to the specific techniques used to implement a monitoring system. We shall now investigate the most important of these.

Four basic methods of condition monitoring exist [1]:

(i) Visual Monitoring
(ii) Performance Monitoring
(iii) Vibration Monitoring
(iv) Wear Debris Monitoring

We shall review each in turn and consider the principle techniques used in each category.

1.2.1 Visual Monitoring

Components may be inspected visually to determine their condition without stripping-down the machine. This is the most limited condition monitoring method and depends heavily upon the skill of a human observer. Specific techniques include use of microscopes, boroscopes, stroboscopes, crack dye penetrants and radiographs. Such techniques have been used for many years and do not require significant levels of technological back-up.

1.2.2 Performance Monitoring

Performance monitoring allows machine condition to be estimated through monitoring transfer functions. Specific wear processes are indicated by various effects on machine performance. Performance monitoring is a relatively new method and is yet to be established in industry.

1.2.3 Vibration Monitoring

Vibration monitoring measurements are used to detect wear conditions which produce changes in a machine's vibrational spectrum. Systems for vibration monitoring usually comprise a vibration transducer and associated electronics in combination with some form of spectral analysis device. Such systems have been commercially available for some time and are recognised by industry as being best employed for condition monitoring of high-speed rotational plant.

1.2.4 Wear Debris Monitoring

Wear debris monitoring of lubricants is the fourth basic condition monitoring method and enables the condition of lubricant-wetted component surfaces to be estimated. The lubricating oil acts as a transport medium for wear debris, and the oil flow is monitored at some convenient point. In the past off-line spectrometric analysis and magnetic plug techniques have been predominant. Recently, however, more technologically sophisticated techniques, including on-line types, have been introduced. This project is concerned with designing, building and testing a potential on-line system.

1.3 THE OVERALL PROJECT OBJECTIVE AND THE RESEARCH PROGRAMME

The overall objective of the project was to design, build and test a system which could be applied to wear-debris monitoring.

Initially, the subject area was reviewed to obtain an understanding of the area of application. A theoretical study of various possible methods for detecting and identifying wear-debris particles was then made. The resulting overall system plan, consisting of three main components - a transducer, associated drive electronics and subsequent signal processing electronics - was considered viable. Specific project objectives are listed in §2.5.1.

Detailed theoretical studies on the proposed scheme were completed in parallel with the construction of a prototype system which was calibrated and tested using particles of various materials. The prototype system has successfully demonstrated the viability of the theoretical predictions.

1.4 THESIS FORMAT

This thesis is sub-divided into chapters, sections and sub-sections. The equations associated with each section or sub-section are labelled using the name of the section or sub-section followed by the equation number in parenthesis. Figures are labelled similarly, with the section or sub-section name being followed by the figure letter in parenthesis.

There are nine chapters. The fields of condition monitoring and wear-debris monitoring are introduced in Chapter One which ends with a description of general project objectives and thesis format. Chapter Two includes a review of lubricant wear-debris monitoring principles and techniques, both commercially available and in development, and concludes with a list of specific project objectives. The theory of the inductive coil transducer is developed in Chapter Three while Chapter Four deals with the marginal oscillator drive circuitry and Chapter Five is on the subsequent FM and AM demodulation signal processing techniques. Theoretical overall characteristics of the system are obtained in Chapter Six while Chapter Seven describes the system experimental set-up as well as calibration and testing using particles of various materials. Chapter Eight, the discussion Chapter, relates the findings of Chapter Seven to the theory of Chapters Three to Six as well as assessing the potential for applying the system to wear-debris condition monitoring. Finally, in Chapter Nine, conclusions and recommendations for future work are described.

CHAPTER ONE: REFERENCE

1. Neale M & Associates: 'A Guide to the Condition Monitoring of Machinery', HMSO, London (1979).

CHAPTER TWO

WEAR-DEBRIS DETECTION TECHNIQUES FOR LUBRICANT BASED CONDITION MONITORING

This chapter describes the nature of lubricant wear debris and various approaches available for their detection, including those in production and at the research and development stage, are outlined. Particle detection and analysis techniques currently used in other areas are assessed for applicability to this work and matched against ideal objectives.

2.1 LUBRICANT BORNE WEAR DEBRIS CHARACTERISTICS AND THEIR USE FOR CONDITION MONITORING

Wear debris monitoring was introduced in Chapter One and we shall now review, in more detail, the characteristics of the debris and their significance for condition monitoring purposes.

It is widely accepted [1] that, by monitoring the quantity and nature of wear debris, an indication of the condition of the various machine components which are in contact with the lubricant may be obtained. The nature of wear debris is associated, through particle material, size and shape, with the deterioration of specific machine components under different modes of wear.

2.1.1 Lubricant Flow for a Typical Machine

A lubricant flow in a machine may be modelled as shown in Figure 2.1.1(a), and will include filters to clean the lubricant of debris. If the debris we wish to monitor is absorbed by the filter, we must monitor at a point in the flow-line between the machine and the filter. If the debris of interest flows through the filter, then the position of the monitoring point is not critical. In general, however, even fine debris particles will not be certain to pass through the filter and it is usually advisable to monitor at the point shown.

The lubricant can, to a good approximation, be regarded as being an infinitely resistive medium with a value of relative magnetic permeability of unity.

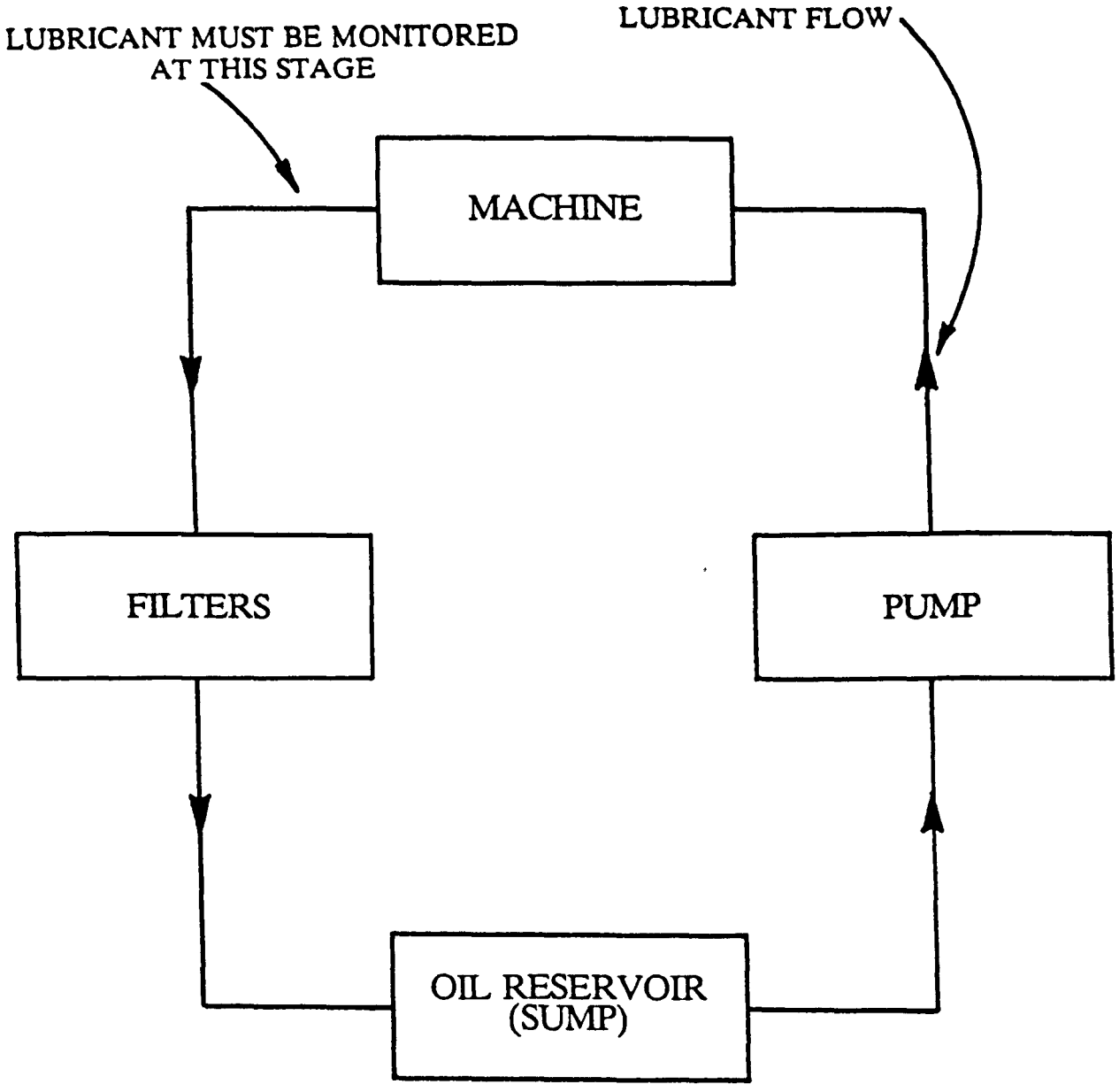


Figure 2.1.1(a) A schematic diagram of the lubricant flow system of a typical machine.

2.1.2 Wear Processes and Associated Wear Particles

Wear particle distributions may be characterised by four key features:

(i) Material
(ii) Size
(iii) Shape
(iv) Concentration

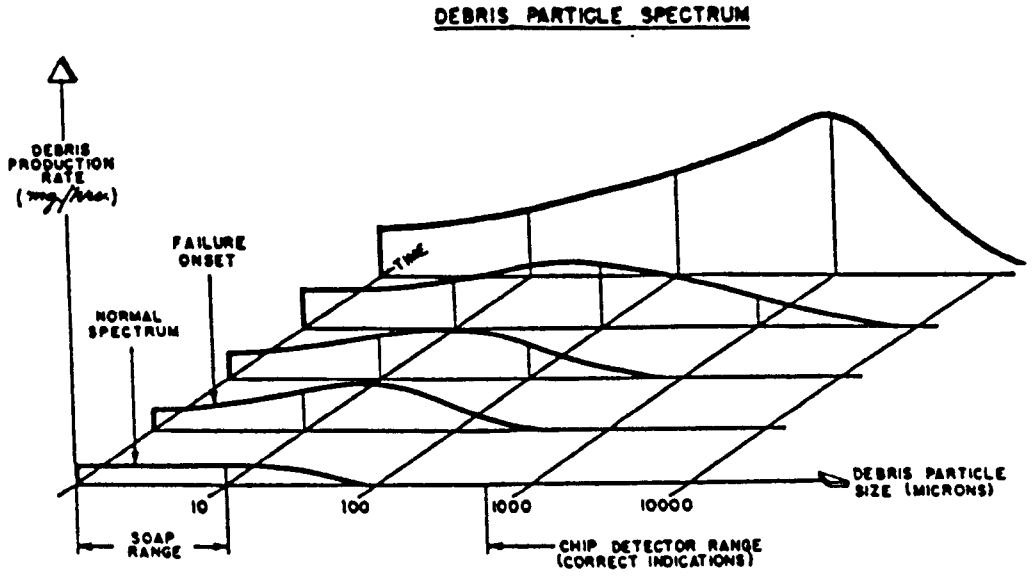
Identification of some or all of these features enables specific wear processes occurring in corresponding components in a machine to be estimated. Listings of machinery problems commonly indicated by lubricant debris content are available, although often contradictory.

Probably the most detailed listing is to be found in the Government sponsored report by Neal [1] which gives correlations between the first three of the above features and various wear processes. The other major listing, by Bowen et al [2], is less comprehensive and is aimed specifically for use in conjunction with ferrographic equipment. In addition, the introductory sections of several papers on lubricant monitoring techniques [3,4,5,6,7] give an indication of expected particles. Not surprisingly, however, they tend to emphasise the significance of those particles best detected by the techniques being discussed. It is consequently difficult to be specific about the relative importance of different debris distributions although some general conclusions can be drawn.

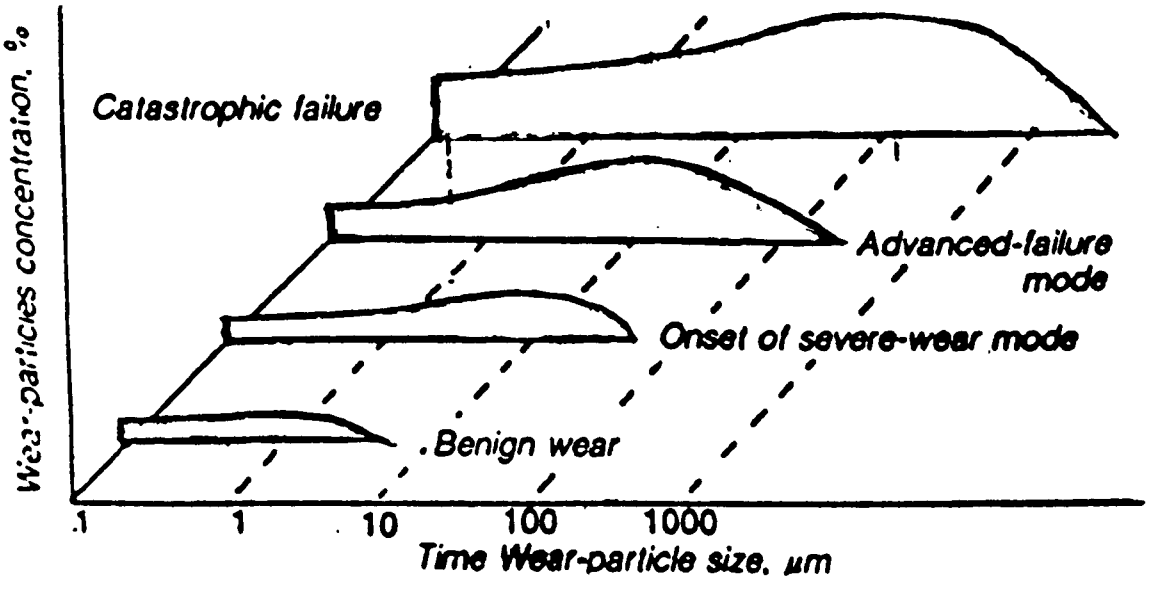
The first and universally agreed principle of debris analysis is that the higher the concentration of metal particles and the larger the particles being generated, the more severe the wear. However, a general definition of significant wear particle concentrations seems unavailable although serious wear has been correlated with increased levels of particle concentration for specific items of plant (e.g. for diesel engines [5,8]).

Tauber [9] shows that normal wear processes generate some particles up to 100μm diameter although most debris comprises particles of less than 10μm. As a component begins to fail, the distribution of particles changes such that the total debris content increases and the amount in the form of larger particles becomes proportionally greater. A remarkably similar diagram is presented by Baur [10], in a paper on ferrography (see §2.2.3), although he has rescaled his particle-size axis to show normal wear generating particles always less than 10μm diameter. These diagrams are reproduced as Figure 2.1.2(a).

Whilst authors reporting on ferrographic techniques indicate particles as small as 5μm being of significance for warning of abnormal wear, papers dealing with chip



(i)



(ii)

Figure 2.1.2(a) Contrasting diagrams showing the particle size spectra as wear progresses in a hypothetical machine; (i) due to Tauber [9], and (ii) due to Baur [10].

detectors of various types claim particles of 50μm and above being important. It is possible that small particle detection gives earlier warning of an incipient failure than monitoring larger particles. Ideally, we would require information on particles ranging from 1μm to 1mm in diameter. (Note that throughout this work the term 'particle size' refers to particle diameter, or approximate diameter for non-spherical particles.)

The second major principle of debris analysis is that we should be able to identify the particular component from wear particle shape and material. This requires a listing of correlations between component wear and lubricant borne particles. This may be obtained through a combination of historical data for a particular machine and the machine type, as well as direct evidence based on knowledge of component materials.

2.2 LUBRICANT-BASED CONDITION MONITORING TECHNIQUES

Over many years a number of lubricant wear-debris monitoring techniques have been introduced, each having associated advantages and drawbacks. We shall now categorise and review the most well known techniques used by industry.

Wear-debris monitoring may be classified into three main types:

(i) Direct Detection Techniques

(ii) Debris Collection Techniques

(iii) Lubricant Sampling and Analysis

2.2.1 Direct Detection Techniques

The direct detection approach involves arranging the oil flow through a sensor which is sensitive to characteristics of the wear debris. Subsequent electronic signal processing provides an output which can be fed to a visual display and/or to some form of computer for further processing. Direct detection techniques are always on-line and consequently yield a continuous estimate of machine condition. Currently there are several available products in this category.

The Smiths Industries particle detector system detects metal particles through use of inductive coil sensors which comprise part of an ac bridge. Signal processing of pulses generated by the passage of particles activates counter displays which allow assessment of particle concentrations in two size ranges. Individual ferrous particles above ~200μm are detectable with sensitivity to non-ferrous particles being much lower. Recent development work has, however, improved the performance of this device.

Some optical methods also fall into the direct detection group, including the HIAC [11], Environment|One [12], and L & N Microtrac [13] systems. These are all based on measuring light scattering and attenuation across a sample cell. The opacity of some lubricants prevents universal application of these systems, although the manufacturers' data fails to address this problem.

Other direct detection techniques include electrically conducting filters, where electrically conducting debris is detected by being short circuited across elements of the filter mesh, but these are not in widespread use. Also, capacitive techniques have been developed but these too have failed to gain general acceptance by industry.

2.2.2 Debris Collection Techniques

The second category of wear-debris monitoring, debris collection techniques, involve collecting debris continuously in a device which is periodically removed for examination. An immediate indication of debris concentration may be obtained by an inspection of the collection device, with more detailed information on particle material and size distribution possible after laboratory analysis. Debris collection techniques are relatively inexpensive and well tried, but suffer from some important drawbacks. An estimate of machine condition cannot be known continuously, and frequent inspection of the collection device may be necessary to avoid machine breakdowns. Each inspection will require the costly use of skilled personnel.

The most common product available for debris collection is the magnetic plug which is fixed in the wall of a lubricant flow-line and attracts ferrous debris. These have been in use for many years and suffer from the disadvantages listed above, as well as an inability to detect non-ferrous particles.

Normal system filters may be viewed as a debris collection technique and the filter sludge can be analysed for wear particles. The main problems with this approach are that most machines have to be stopped for filter removal, the oil system may have to be drained prior to filter removal, and filters generally do not allow easy display or removal of wear particles. However, special filter units are available with mechanisms which collect debris in a removable chamber. Such filters feature a capability to capture non-ferrous debris and are often used in conjunction with magnetic plugs for this reason.

2.2.3 Lubricant Sampling and Analysis

Lubricant sampling and analysis involves extracting a sample of lubricant from the machine and subsequently analysing the sample for wear debris. It is best to obtain the sample whilst the machine is running or shortly after stopping to ensure particles are still in suspension. The analysis can take two forms:

(i) To determine the concentration of chemical elements.

(ii) To determine the size, concentration and shape of particles.

These approaches are complementary in that, provided the debris is sufficiently fine, the former does not depend on the form of the debris but only on the *elementary* concentration, whilst the latter approach allows only very limited material determination.

The main elemental analysis techniques [1] are atomic absorption, atomic emission and X-ray fluorescence. Atomic absorption operates on the principle that every atom absorbs light only of its own characteristic frequency. After dilution, the oil sample is vapourised in a flame. A direct measure of the concentration of each element is then obtained by noting the light energy absorbed when the element's characteristic frequency is shone through the flame. Most metals can be detected using such a system, and it is the least expensive of the elemental analysis techniques. However, it is practicable to test relatively few samples and for a limited number of elements, say 25 samples for four elements per man day, and only particles less than $\sim 15\mu m$ can be analysed.

Atomic emission spectrometers depend on elements present in the oil sample emitting characteristic wavelengths of light when the sample is excited by, say, an electric discharge. This technique detects larger particles than absorption techniques and is significantly faster and more comprehensive. One system performs a test in less than 60 seconds and gives concentrations of 25 elements. Emission spectrometric techniques have been available on a contract basis for many years where samples are sent to a specialist laboratory by machine operators and results of elemental concentrations returned. In addition, advice on changes and trends of concentrations may also be obtained.

X-ray fluorescence is the process where elements emit characteristic x-ray frequencies on being exposed to a radiation source which is usually radioactive.

Detection of most metals is possible but high equipment cost has prevented widespread use of this technique.

A significant drawback of the elemental techniques is an inability to identify easily the materials of particles of alloys or oxides. Also, particles greater than ~20μm do not respond well to atomic absorption and emission techniques, and this limits their effectiveness.

The other category of lubricant sampling and analysis techniques comprises those which produce information on particle shape, size and concentration. Microscope-based analysis of wear debris can be useful, but ferrographic methods [14] have been introduced over the last ten years and these allow more detailed investigation.

Ferrography involves running the oil sample over a glass slide which is, in turn, suspended over a strong permanent magnet, arranged to produce a magnetic field which increases in strength along the length of the slide. Ferromagnetic debris is deposited on the slide in order of size, and the particles are fixed using chemical agents. The slide can then be analysed under a microscope specially illuminated to show up metal particles whose composition may be estimated by observing colour changes due to heating. The operation of most ferrographic equipment requires highly skilled personnel, but samples may be sent to specialist laboratories for analysis. Although the equipment is limited to detecting ferromagnetic particles a relatively large range of particle sizes, from a few μm to about 1mm, can be dealt with.

2.3 CURRENT RESEARCH IN LUBRICANT DEBRIS MONITORING TECHNIQUES

Over recent years interest in condition monitoring has encouraged research into improved wear debris monitoring techniques, particularly those which operate on-line. In this section we shall review the most well-known of such techniques.

2.3.1 The Smiths Industries Metal Particle Detector

Research and development of an improved metal particle detector system has been undertaken by Smiths Industries during the last few years [15], the new system being a refined version of that described in §2.2.1. Based on the same inductive-sensor principle, the new system features improved sensitivity, which is claimed to allow detection of iron particles as small as 100 μm in diameter. The improvement has been achieved mainly through mounting signal condition electronics directly on the sensor. Modern microprocessor techniques have also been applied for manipulating data on trends in

particle size distribution changes.

It is worth noting that although discrimination between ferrous and non-ferrous metal particles has been achieved, the particular metal cannot be determined.

2.3.2 An Ultrasonic Monitoring Device

Another current area of research for lubricant debris monitoring is the use of ultrasonic scattering, and this is being investigated by Harries et al [16]. Their system is based on the principle that debris particles scatter an ultrasonic signal. The transducer assembly consists of an ultrasound transmitter and adjacent receiver. The positioning of the transducers and the frequency of the ultrasound are critical, and this may make industrial application of this system difficult. The vibration in an operational machine would probably cause severe performance degradation; the ability of the system to detect particles of 35μm, which is claimed for the experimental set-up, may well be swamped by anomalous vibrational effects.

The use of this ultrasonic technique permits detection of relatively hard particle of any material, which cannot be achieved using electromagnetic field techniques. However, identification of specific particle materials cannot be made and certain lubricating oils may cause severe ultrasonic attenuation [17], resulting in poorer sensitivity.

2.4 THE POTENTIAL FOR USING CURRENTLY AVAILABLE PARTICLE DETECTION & ANALYSIS TECHNIQUES

A range of particle detection and analysis techniques is described in the literature and the applicability of these to lubricant monitoring is reviewed here. We have studied two main information sources: general review papers and more specific research papers.

Two major review papers appear in the literature: one by Hinde [18] and the other by Stanley-Wood [19]. Hinde's paper provides a review of real-time particle size analysers with a strong bias towards the mining and metallurgical industries. He describes a range of techniques including measurements on slurry viscosity, sieving and other mechanical methods, as well as various optical and ultrasonic techniques but none appears particularly suitable for lubricant monitoring purposes. Perhaps surprisingly, there is no mention of methods based on electromagnetic principles.

In his review, Stanley-Wood discusses a range of 'on-stream' particle size and size distribution measurement instruments and includes some techniques which have been used for lubricant monitoring (e.g. the HIAC particle counter). The scope of this

article is illustrated by the block diagram reproduced as Figure 2.4(a). Another review, by Larsen et al [20], covers roughly the same area as Stanley-Wood's but in much less depth.

Numerous specialised papers have appeared over the last twenty years offering a range of particle analysis techniques and these may be grouped according to their physical principles. Electric field based techniques, such as those dependent upon electrostatic separation effects [21], optical property changes due to an applied electric field [22] and spark discharge effects [23] have been reported. None of these seem suitable for lubricant monitoring.

Coulter counter (or liquid resistor) techniques have been in use for many years and these work on the principle that particles flowing through a calibrated narrow-orifice resistance cell modulate the resistance. This technique is only effective when the suspending medium exhibits a measurable resistivity whilst the particles of interest have a considerably different resistivity. Such circumstances occur in many biological applications [24] but the extremely high resistivity of lubricating oil makes such a technique inappropriate for our purposes.

A number of optical particle analysis techniques has been developed and these are predominantly laser based. General systems capable of detecting particle distributions using the principle of light scattering are widely reported [25,26,27] whilst other authors describe systems aimed at specific applications and these fall into three main groups: for monitoring atmospheric particles [28,29], for process control applications [30,31,32] and for blood cell analysis [33,34]. The variability and high degree of opacity of lubricating oil prevents the general use of any of these optical methods although, in principle, some of them could be applied to machinery where relatively transparent oils are used.

None of the methods described in this section appears to offer particularly good potential for application to lubricant monitoring. It is the nature of lubricating oil in general, with its high resistivity and opacity, which presents the main problem.

2.5 IDEAL LUBRICANT MONITORING TECHNIQUES AND PROJECT OBJECTIVES

Having reviewed condition monitoring in general and current lubricating techniques in particular, and having also looked into the possibility of using currently available particle detection and analysis techniques, it is now appropriate to identify the characteristics of an ideal lubricant monitoring system. This leads directly to an outline of

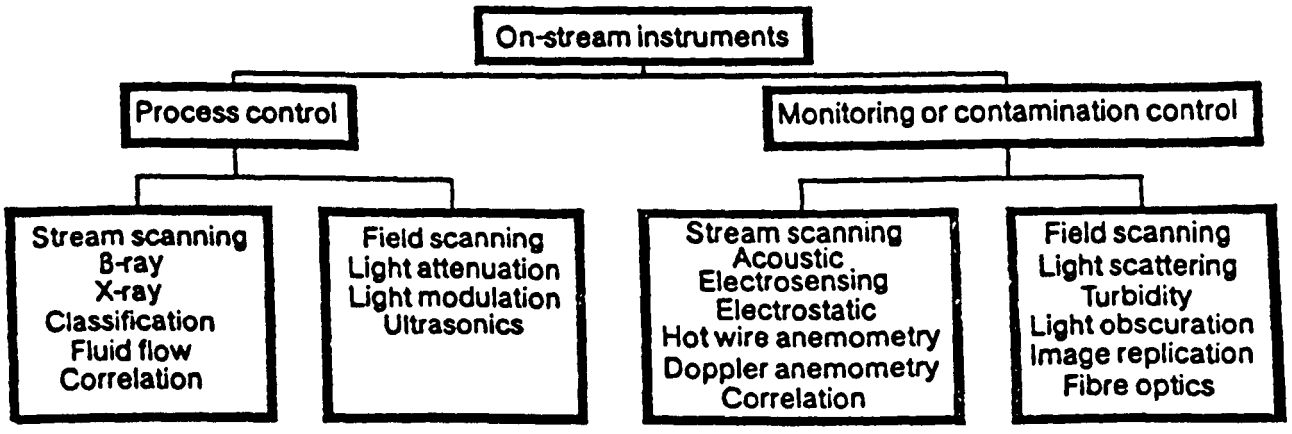


Figure 2.4(a) The range of techniques for 'on-stream' particle size and size distribution as described by Stanley-Wood [19].

specific project objectives.

2.5.1 An Ideal Lubricant Monitoring System

The definition of an ideal lubricant monitoring system will vary according to the specific application but the following features would probably be included:

(i) On Line.

(ii) Low capital and maintenance cost relative to the cost of the machinery being monitored and its associated maintenance programme.

(iii) Easy to install, either as part of the production of the monitored machinery or as a retrofit.

(iv) Able to detect particles in the approximate size range 1μm to 1mm.

(v) Able to to identify particle shape, size, composition and concentration.

The above combination is, in reality, impossible to achieve at present. Indeed, currently available systems (see §2.2) do not even approach the ideal. A compromise must be obtained, with low cost normally being obtained in exchange for a level of performance considerably lower than the ideal. Even where cost is not a limitation, no technique currently available or in development would satisfy the remaining criteria. There is considerable scope for research in this area and arguably any significant addition to the techniques presented in §2.2 and §2.3 should be valuable.

2.5.2 Specific Project Objectives

As described above, the development of an ideal lubricant monitoring system is probably not possible with the level of technology available today. Worthwhile improvements are, however, still attainable and with this in mind we compiled a list of project objectives. These were to build and test a system which would:

(i) Operate on-line using an electromagnetic transducer technique to detect individual particles, particularly of metals, e.g. brass and steel.

(ii) Allow detection of particles in the ranges specified as significant by previous workers.

(iii) Produce output signals from which an indication of particle size and composition could be easily obtained.

(iv) Be cheap and rely on sophisticated electronics rather than a complex mechanical transducer.

We decided to use some form of electromagnetic transducer to circumvent problems associated with optical, ultrasonic and mechanical techniques (see previous sections) and we were also encouraged by the success of manufacturers who had previously used electromagnetic techniques for lubricant monitoring. We felt that detection of metal particles with diameters as small as 100μm should be possible. In addition, analogue circuits could be constructed at fairly low cost and we considered that their use in conjunction with an electromagnetic transducer, could provide output signals suitable for connection, in subsequent work, to analogue to digital conversion and computer systems.

2.6 SUMMARY AND CONCLUSIONS FOR CHAPTER TWO

This chapter is summarised and concluded as follows:

- The characteristics of wear-debris in the lubricant flow of a machine can indicate the nature and severity of wear in oil-wetted components.
- The greater the severity of the wear, the higher the concentration and size of the wear-debris particles. Depending on the reference cited, particles greater than threshold values in the range 10μm to 100μm are indicative of severe wear.
- Most current wear-debris monitoring techniques operate off-line.
- Recent research into wear-debris monitoring has concentrated on on-line techniques. Systems based on both inductive sensors and ultrasonic sensors have been investigated.
- In general, particle detection and analysis techniques currently applied in other fields are not easily applicable to wear-debris monitoring.
- An ideal wear-debris monitoring system would be on-line, low cost, easily installed and able to detect particles in the size range 1μm to 1mm. Moreover, it would be capable of identifying particle shape, composition and concentration as well as size. Such an ideal system is not, however, feasible.
- We identified specific project objectives which consisted of designing, building and testing a wear-debris monitoring system which would detect individual metal particles (e.g. brass, steel) as small as 100μm, produce output signals from which an indication of particle size and composition could be obtained, be relatively cheap, and involve sophisticated electronics rather than a complex mechanical transducer.

- The use of an electromagnetic transducer was considered the best option.

CHAPTER TWO: REFERENCES

1. Neale M and Associates: 'A guide to the condition monitoring of machinery', HMSO, London (1979).

2. Bowen E R and Wescott V C: 'Wear particle atlas Volume 1', Foxboro Analytical, Burlington Mass USA, c1980 (undated).

3. Hunter R C: 'Engine failure prediction techniques', Aircraft Engineering, 47 (1975) 4-14.

4. Barwell F T: 'Particle examination - a tool for maintenance engineers', Chartered Mechanical Engineer, (1978) 33-36.

5. Evans C: 'Wear debris analysis and condition monitoring', NDT International, (1978) 132-34.

6. Scott D: 'Condition monitoring', Tribology International, (1978) 29-33.

7. Scott D, Seifert W W and Wescott V C: 'The particles of wear', Scient Amer, **230** (1974) 88-97.

8. Morley G R: 'Comparison of spectrometric and ferrographic analysis of crankcase oils from the high speed train', Tribology International, (1981) 159-65.

9. Tauber T: 'A new chip detector', Aircraft Engineering, **49** (1977) 4-6.

10. Baur P S: 'Ferrography: machinery-wear analysis with a predictable future', Power, **126** (1982) 114-117.

11. ---: 'The HIAC PC-320 (systems to size and count particles)', data sheets, HIAC Instruments Division, Montclair CA USA, August 1981.

12. ---: 'Generator condition monitor', data sheet, Environment | one corporation, Schenectady NY USA, May 1975.

13. ---: 'L & N Microtrac Suspended Solids Monitor', data sheet, Leeds & Northrup, North Wales PA USA, 1979.

14. Hampson L G: 'Industrial uses of ferrography', Chartered Mechanical Engineer, May 1981, 60-62.

15. Masom R A: 'The development, proving and application of an in-line metal particle detector (MPD)', Brit J NDT, May 1985, 159-166.

16. Harries C J, Sayles R S & Macpherson P B: 'An on-line ultrasonic device for monitoring contamination in fluids', 5th National Conference on Condition Monitoring, U.K., March 1984, 660-672.

17. Heyman J S, Dietz D and Miller J G: 'A non-doppler ultrasonic monitor for particulates in flowing fluids', IEEE Ultrasonic Symposium Proceedings (1975) 561-564.

18. Hinde A L: 'A review of real-time particle size analysers', J South African Inst Mining and Metallurgy, **73** (1973) 258-268.

19. Stanley-Wood N G: 'Measuring on-stream particle size and size distribution', Control & Instrumentation, (December 1974) 42-47 and (January 1975) 30-35.

20. Larsen P S, Engelund F, Sumer B M and Lading L: 'Techniques of sizing and tracking of particles', J Fluid Mech, **99** (1980) 641-653.

21. Inculet I I, Murata Y and Castle G S P: 'A new electrostatic separator and sizer for small particles', IEEE Trans Indust Appl, **IA-19** (1983) 318-323.

22. Oakley D M, Jennings B R, Waterman D R and Fairey R C: 'An electro optic birefringence fine-particle sizer', J Phys E: Sci Instrum, **15** (1982) 1077-1082.

23. Auer S and Berg O E: 'Composition analyser for microparticles using a spark ion source', Rev Sci Instrum, **46** (1975) 1530-1534.

24. Merrill J T, Veizades N, Hulet H R, Wolf P L and Herzenberg L A: 'An improved cell volume analyser', Rev Sci Instrum, **42** (1971) 1157-1163.

25. Suda K: 'Instrumentation for the size determination of submicron particulates systems by sideways light scattering method', Rev Sci Instrum, **51** (1980) 1049-1058.

26. Durst F: 'Combined measurements of particle velocities, size, distributions and concentrations', Trans ASME, **104** (1982) 284-296.

27. Yamauchi T and Ohyama Y: 'A study on the measurement of particle size distribution with laser diffraction systems', Bull JSME, **25** (1982) 1931-1937.

28. Reagan J A, Byrne D M and Herman B M: 'Bistatic LIDAR: a tool for characterising atmospheric particulates: part 1 - the remote sensing problem', IEEE Trans Geosc & Rem Sensing, **GE-20** (1982) 229-243.

29. Knollenberg R G and Gilland J R: 'Pioneer Venus sounder probe particle size spectrometer', IEEE Trans Geosc & Rem Sensing, **GE-18** (1980) 100-104.

30. Llewellyn G J: 'A microprocessor-based analyser for data derived from slurry flows for in-stream particle size determination', J Microcomputer Applications, 6 (1983) 25-35.

31. Wertheimer A L, Frock H N and Muly E C: 'Light scattering instrumentation for particulate measurements in processes', SPIE, **129** (1977) 49-58.

32. Muly E C, Frock N H and Weiss E L: 'The application of Fourier imaging systems to fine particles', J Powder & Bulk Solids Technology, **2** (1978) 3-8.

33. Cornillault J: 'Particle-size analyser', First Electro-optics Markets & Technology Conference, Geneva (1972) 289-288.

34. Miller M N: 'Design and clinical results of Hemtrac; an automated differential counter', IEEE Trans Biomed Engrg, **BME-23** (1976) 400-405.

35. Shi N K and Carlson F P: 'Application of binary filters in automated counting of reticulocytes', IEEE Trans Biomed Engrg, **BME-23** (1976) 84-88.

CHAPTER THREE

THE INDUCTIVE COIL TRANSDUCER

This chapter begins with a review of possible types of electromagnetic transducer for lubricant monitoring applications. The effect of a metal particle on a magnetic field is then considered in preparation for the final sections on the inductive coil transducer which was adopted in this work.

3.1 CHARACTERISTICS OF ELECTROMAGNETIC TRANSDUCERS

Electromagnetic transducers may be divided into three main categories:

(i) Electric Field Dominated (Capacitive).
(ii) Magnetic Field Dominated (Inductive).
(iii) Combination Electric/Magnetic Field Based Transducers.

We shall review each in turn, relating their characteristics to lubricant debris applications.

3.1.1 Electric Field Dominated Transducers

Electric field based, or capacitive, transducers depend on the sample interacting with the applied field to form its own scattering field. Detection of the scattering field gives information on the sample characteristics.

For our purposes we require a relatively small sensing volume (the region in which the particle is to be detected) - say 1 cm^3 - and a capacitive transducer of this size has a capacitance of only a few pF. This capacitance is dominated by the parasitic effects, primarily of connection cabling, which make reasonably sensitive measurements difficult. This problem has been described in more detail by Lim [1] who attempted particle detection with a capacitive sensor wrapped around a cylindrical glass pipe. An advantage of the electric field transducer is its ability to detect non-metal particles but this is outweighed by lack of sensitivity for lubricant monitoring purposes.

3.1.2 Magnetic Field Dominated Transducers

Magnetic field dominated transducers involve subjecting the sample volume to a magnetic field and monitoring the resulting scattering field. This process can be achieved through use of separate exciting and detection coils, although better sensitivity is obtained through use of a single coil for both purposes.

The inductive coil transducer, for a sensing volume of about $1cm^3$, is not dominated by parasitic inductances and this makes measurements relatively straightforward. Moreover, although we are limited to the detection of metallic particles, these are the wear particles most worthy of monitoring.

3.1.3 Combination Electric/Magnetic Transducers

The combination electric/magnetic field transducers take the form of resonant cavities constructed around the sampling volume. They operate at microwave frequencies and perturbations in the quality factor, Q, and resonant frequency yield information on the contents of the enclosed sensing volume.

The cavity transducer has serious drawbacks for use in an on-line metal particle detector. The frequency of operation would require expensive and physically large signal generators which may involve a safety hazard. The sensing volume would have to be restricted to a particular position within the cavity, as the fields of both types vary greatly. This would result in a reduction in sensitivity. In addition, to maximise the information obtained, we require to operate with fields inside the particle partially skin-depth limited (see §3.2) and microwave frequencies would result in severe skin depth limiting.

3.2 THE INTERACTION OF METAL PARTICLES AND THE MAGNETIC FIELD

In this section consideration is given to the effect on the magnetic vector potential due to the introduction of a spherical metal particle into a previously uniform alternating magnetic field. This analysis provides the basis for evaluating the effect of such a particle on a coil transducer.

3.2.1. The Spherical Metal Particle in the Magnetic Field

Assume that in the region of space of interest we have a uniform alternating magnetic field of angular frequency ω. Referring to Figure 3.2.1(a), and using polar coordinates, this field can be described by:

$$\mathbf{B}_i = B_z\mathbf{a}_z = B_z\cos\theta\ \mathbf{a}_r - B_z\sin\theta\ \mathbf{a}_\theta \tag{1}$$

or, equivalently by:

$$\mathbf{A}_i = \frac{1}{2}rB_z\sin\theta\ \mathbf{a}_\phi \tag{2}$$

from which (1) can be obtained by substitution in $\mathbf{B} = \nabla \times \mathbf{A}$.

Now, if a metal sphere of radius a, conductivity σ and magnetic permeability μ is introduced into the original field, the net vector potential outside the sphere is given by:

$$\mathbf{A} = \mathbf{A}_i + \mathbf{A}_{sc} \tag{3}$$

where $\mathbf{A}_{sc}$ is the scattering potential due to the particle. From the work of Smythe [2], the scattering potential, obtained by solving Laplace's Equation, is given by:

$$\mathbf{A}_{sc} = [\tfrac{1}{2}B_zD/r^2]\sin\theta\ \mathbf{a}_\phi \tag{4}$$

where D is a complex function of a, ω, σ and μ. Inside the sphere, the net vector potential is given by:

$$\nabla^2\mathbf{A} - j\omega\sigma\mu\mathbf{A} = 0 \tag{5}$$

whose solution has the form:

$$\mathbf{A} = [\tfrac{1}{2}B_zCr^{-\frac{1}{2}}\sin(\theta)I_{3/2}[(j\omega\sigma\mu)^{\frac{1}{2}}r]\ \mathbf{a}_\phi \tag{6}$$

where C is a complex function of a, ω, σ and μ, and $I_{3/2}$ is a modified Bessel function.

By considering the magnetic boundary conditions at the surface of the sphere, the complex functions C and D are obtained:

$$C = \frac{3\mu\nu a^{3/2}}{(\mu - \mu_o)\nu I_{-\frac{1}{2}} + [\mu_o(1+\nu^2) - \mu]I_{\frac{1}{2}}} \tag{7}$$

$$D = \frac{(2\mu + \mu_o)\nu I_{-\frac{1}{2}} - [\mu_o(1+\nu^2) + 2\mu]I_{\frac{1}{2}}}{(\mu - \mu_o)\nu I_{-\frac{1}{2}} + [\mu_o(1+\nu^2) - \mu]I_{\frac{1}{2}}}\ a^3 \tag{8}$$

where $I_n[(j\omega\sigma\mu)^{\frac{1}{2}}a]$ is written I_n and $\nu = (j\omega\sigma\mu)^{\frac{1}{2}}a$

For this work it is the fields external to the sphere which are important and as a consequence, the D-function is the more useful. The related D_n function is defined by:

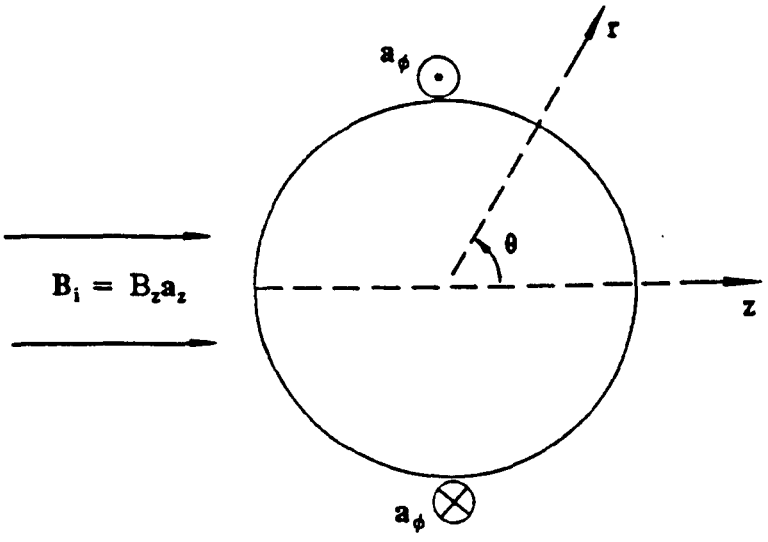


Figure 3.2.1(a) Conducting sphere in a low frequency magnetic field.

$$D_n = D/a^3 \tag{9}$$

Figure 3.2.1(b) shows the real and imaginary components of the D_n function plotted against a/δ, where $\delta = [2/(\omega\sigma\mu)]^{\frac{1}{2}}$ - the skin depth or depth of penetration.

The vector potential at a large distance compared to the radius, a, of a circular current loop which has a current I is given by:

$$A_d = \frac{1}{4}a^2\mu_o I r^{-2}\sin(\theta)a_\phi \tag{10}$$

This is of the same form as equation (4). Equating A_d and A_{sc} gives:

$$\frac{1}{2}B_z D r^{-2}\sin\theta = \frac{1}{4}a^2\mu_o I r^{-2}\sin\theta \tag{11}$$

or

$$\pi a^2 I = 2\pi B_z D/\mu_o \tag{12}$$

Hence the sphere could be replaced by a circular current loop of dipole moment $2\pi B_z D/\mu_o$ to give the same scattering field.

3.2.2. Notable Features of the D_n Function of a Metal Sphere

At extremes of the a/δ ratio, the D_n function tends to simple values, dependent also on the relative permeability. Three situations are of particular interest:

(i) If $a/\delta >> \mu_r$ then $D_n \rightarrow -1$

The behaviour of the sphere tends towards that of a perfect conductor, and currents and fields are confined to a thin outer skin.

(ii) If $\mu_r >> 1$ and $a/\delta < 1$ then $D_n \rightarrow 2\,\frac{\mu_r - 1}{\mu_r + 2}$

The behaviour of the sphere tends towards that in a dc field.

(iii) If $\mu_r = 1$ and $a/\delta << 1$ then $D_n \rightarrow 0$

The effect of the applied field on the sphere becomes negligible as the a/δ ratio tends to zero.

Two distinct current mechanisms contribute towards the D_n function:

1. Conduction currents - i.e. induced eddy currents.

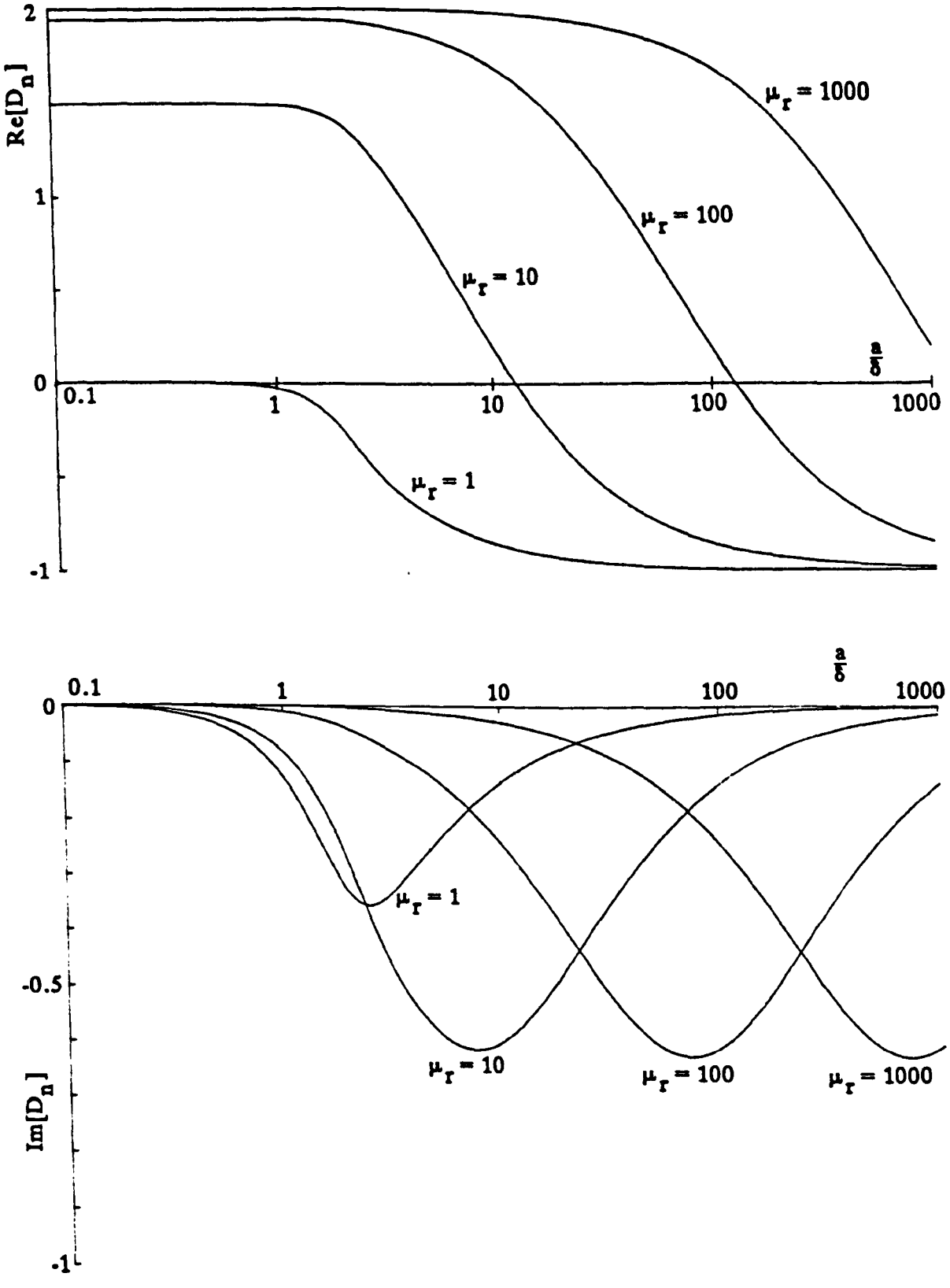


Figure 3.2.1(b) Plots of the real and imaginary components of $D_n = D/a^3$.

2. Magnetisation currents - the effective currents corresponding to the magnetisation effect in magnetic materials.

In the above situations, the D_n function may be considered a function of the relative contributions of these current mechanisms. Maximum information, for particle size and material analysis, can be obtained by operating at conditions intermediate to the extremes listed - i.e. with partial skin depth limiting.

3.2.3 The Relationship between the Volume Magnetic Susceptibility and the D_n Function for a Metal Sphere

The D_n function has, so far, been used to describe the effect on a sphere. As an alternative to the D_n function we may use the volume magnetic susceptibility.

For a metal sphere of radius a, the volume magnetic susceptibility is defined as:

$$\chi = \frac{\pi a^2 I}{H_z(4/3)\pi a^3} \tag{1}$$

where I is the current flowing around the effective current loop as described in §3.2.1. Substituting from Equation 3.2.1(12) yields:

$$\chi = (3/2)D_n \tag{2}$$

or

$$\mathrm{Re}[\chi] = (3/2)\mathrm{Re}[D_n] \quad \text{and} \quad \mathrm{Im}[\chi] = (3/2)\mathrm{Im}[D_n] \tag{3}$$

Clearly, the χ function and the D_n function are the same except for a scaling factor. Indeed, Wejgaard and Tomar [3], whilst restricting their analysis to the unity μ_r case, quote a χ function which produces curves identical to those of Figure 3.2.1(b).

3.2.4 The D_n Function for Specific Materials

Figure 3.2.1(b) shows a set of D_n function curves, plotted against a/δ, for various μ_r values. In this section we consider the D_n and D functions for spheres of certain materials and sizes, with various frequencies of applied field. This is central to the choice of operating frequency. We have chosen copper, brass and lead as representative of non-ferrous metals whilst ferrite and two types of steel represent ferrous materials. Their values of resistivity, ρ, and relative magnetic permeability, μ_r, are listed in §7.3.2.

At a given frequency, the D_n locus follows a path dependent solely on the value of μ_r for the particle. However, the position on the locus, for a given size of particle, depends on the value of its conductivity, σ. Thus, non-ferrous particles, with a value of unity for μ_r, follow the same D_n locus, although the position for a particular size depends on the value of σ. Similarly, the locus paths for ferrous materials depend on the value of μ_r, whilst the position on the locus for a given particle size is controlled by the value of σ for the particle. A family of loci, for various μ_r values is shown in Figure 3.2.4(a).

We need now to consider the D function and its relationship to real particles at specific frequencies. First, note Figure 3.2.4(b), where the locus Re[D] versus Im[D], using particle diameter as the parametric variable, is plotted for particles of different sizes and materials at five different frequencies in the range 100kHz to 1GHz. There are two main points to note. First, the different non-ferrous metals take different locus paths and given the Re[D] and Im[D] components for a particular particle, plotting on an appropriate Re[D]-Im[D] diagram yields a good indication of particle material and size.

Second, the Re[D] versus Im[D] loci for ferrous materials follow paths dependent upon the values of both σ and μ_r. This could cause some ambiguity, as two particles with very different values of σ, μ_r and diameter could generate identical Re[D] and Im[D] ordinates. However, given a series of particles, of a certain material, in different sizes, we could identify the locus curve corresponding to that material. Thus, provided that we do not rely on a single particle to determine the material in question, we should be able to avoid such ambiguity. Figure 3.2.4(b) includes the loci for three ferrous materials and these fall into the quadrant of the D-plane where $Re[D] > 0$, at the lower frequencies for particle sizes of interest.

The best spread of loci curves, for the materials of interest, was obtained at ~10MHz and it was decided to operate at this frequency for this work.

3.2.5 Particle Shape and the D_n Function

The effect of an applied magnetic field on a spherical particle has been reviewed, but, in practice, debris is rarely spherical. However, work from related fields enables us to evaluate the significance of particle shape on the value of the D_n function.

Zimmerman [4] describes the effect of an infinite cylindrical sample in an infinite coil and points out, on a per unit length basis, that the effect differs from that of a spherical sample by a factor of two. In addition, he claims that shape does not appear

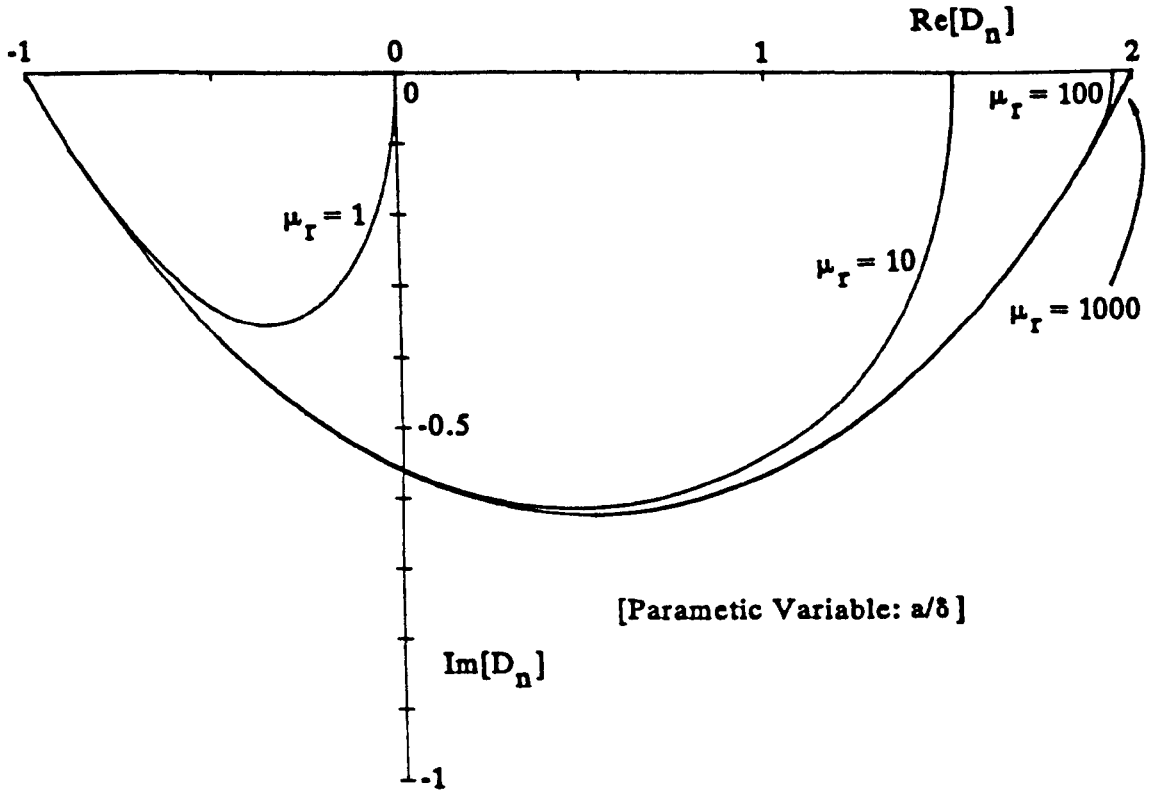


Figure 3.2.4(a) A family of loci for various μ_r values.

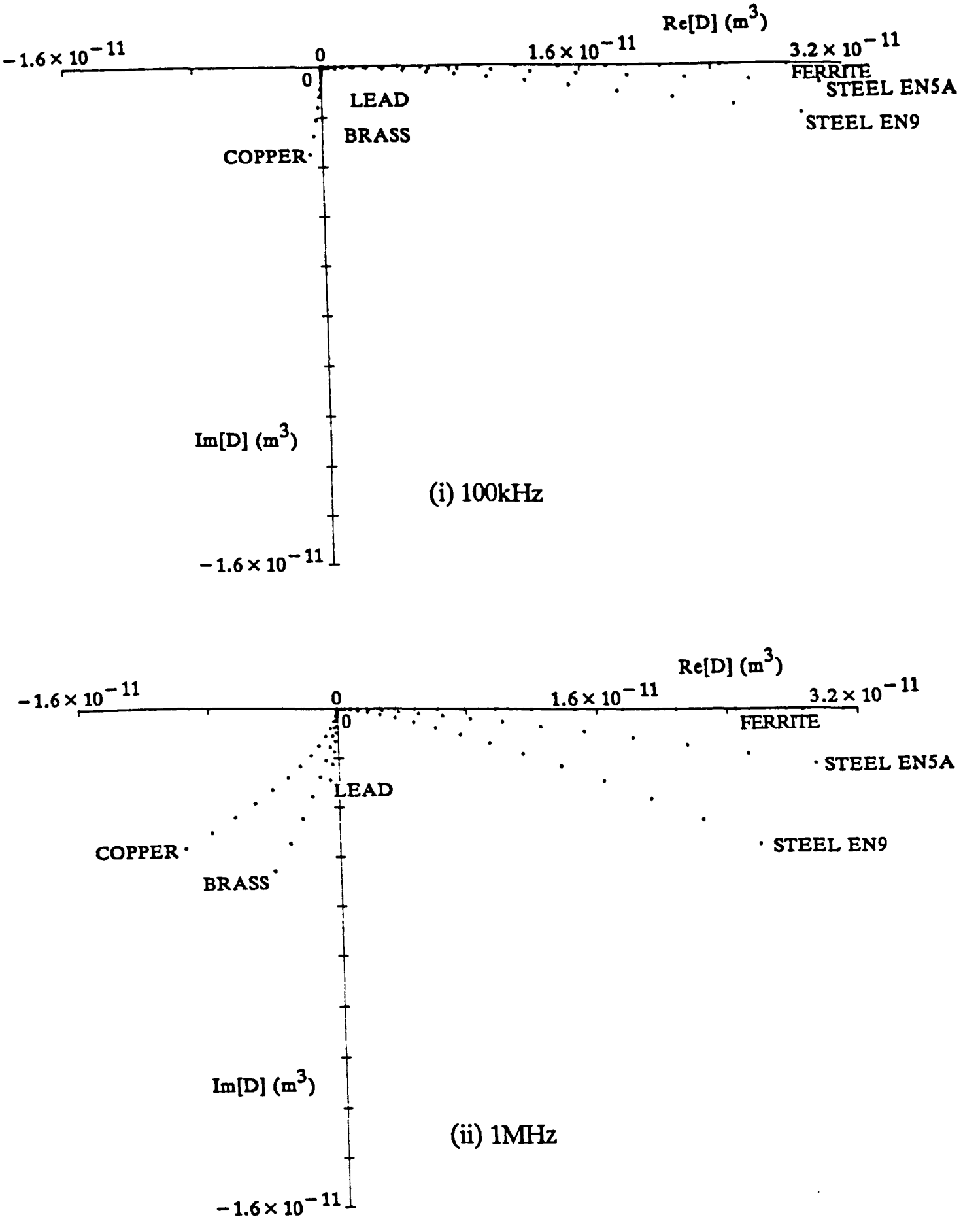
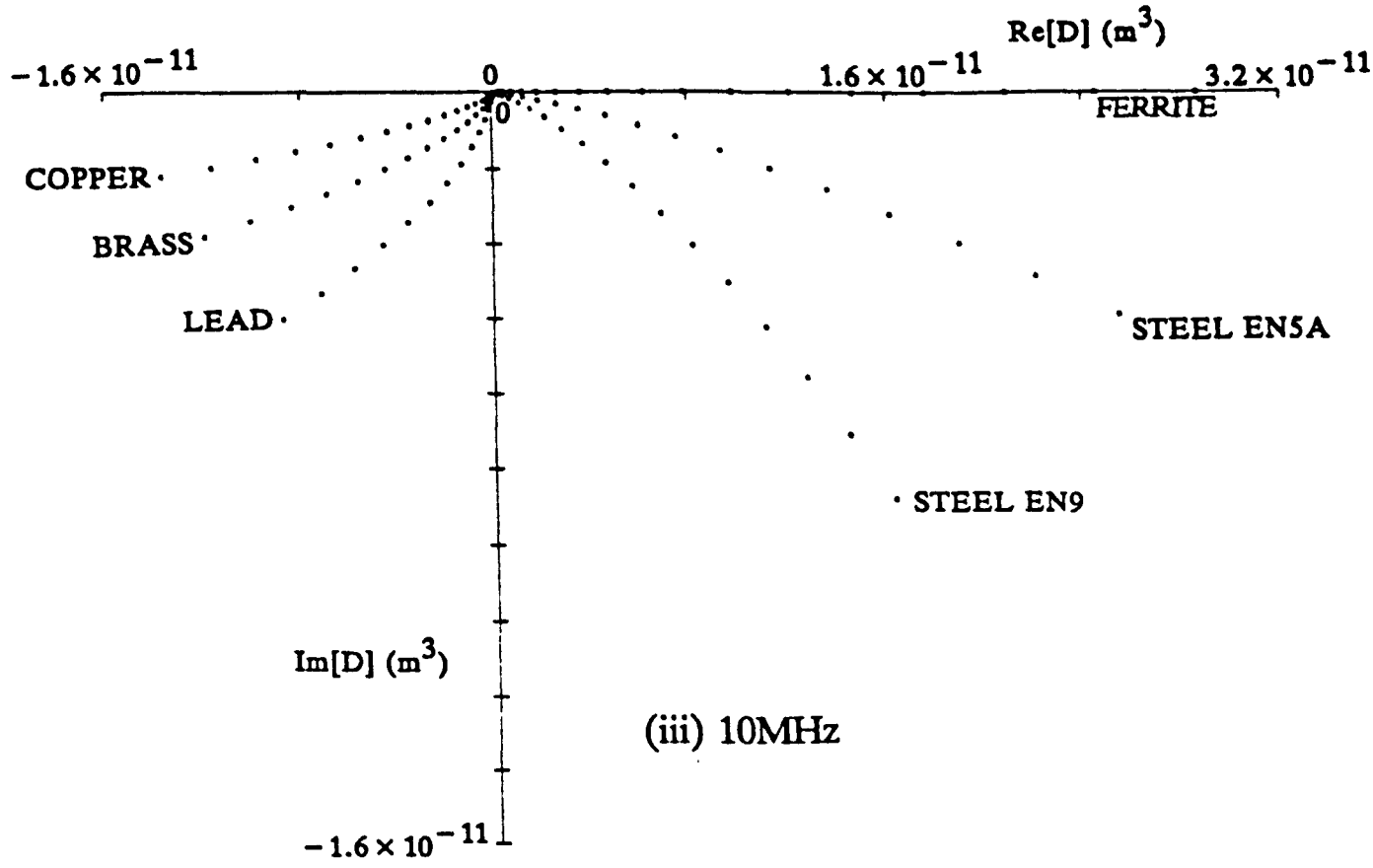
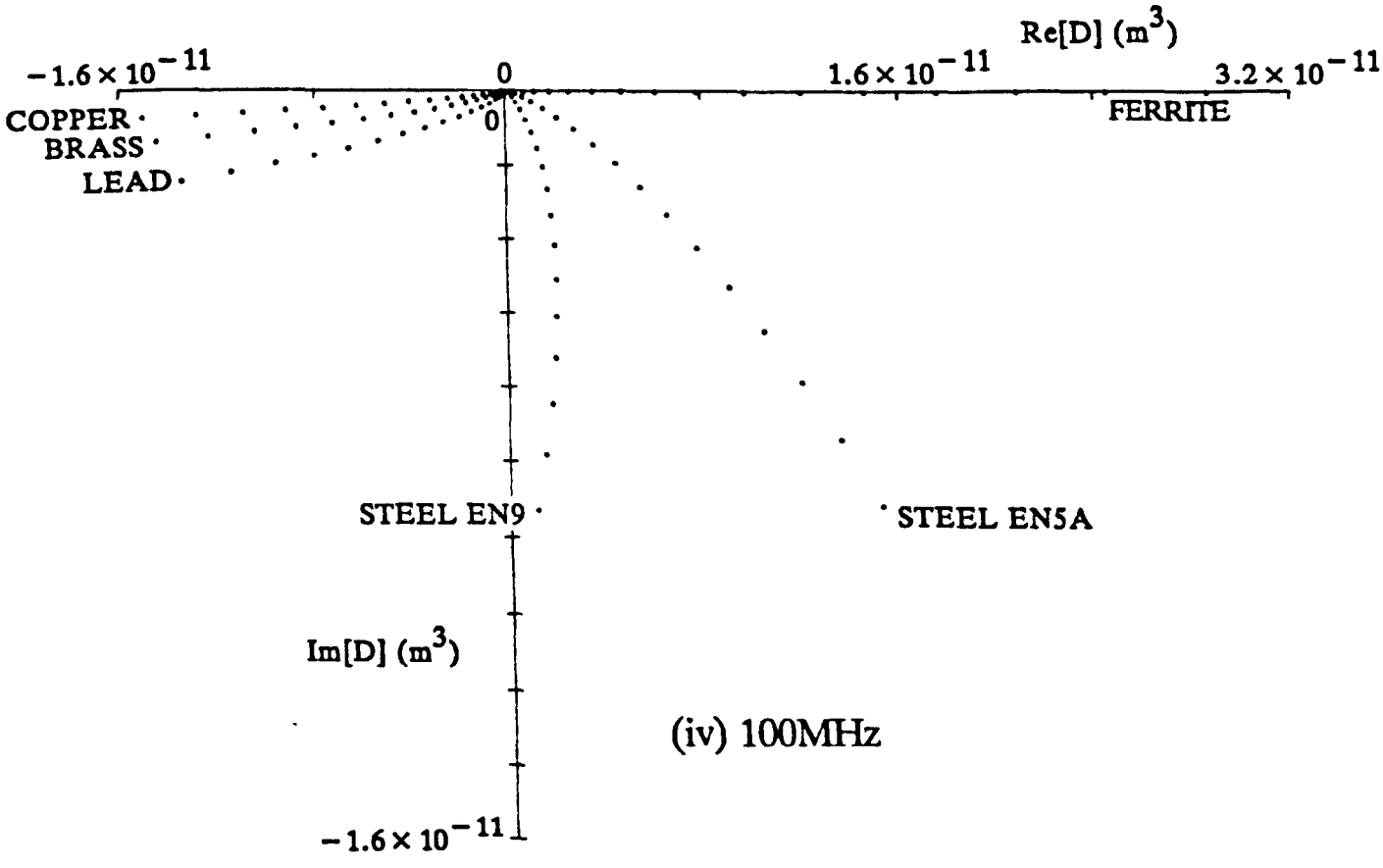


Figure 3.2.4(b) Theoretical D plots for particles up to 500μm diameter at 25μm intervals at frequencies (i) 100kHz, (ii) 1MHz, (iii) 10MHz, (iv) 100MHz and (v) 1GHz.



(iii) 10MHz



(iv) 100MHz

Figure 3.2.4(b) (Continued)

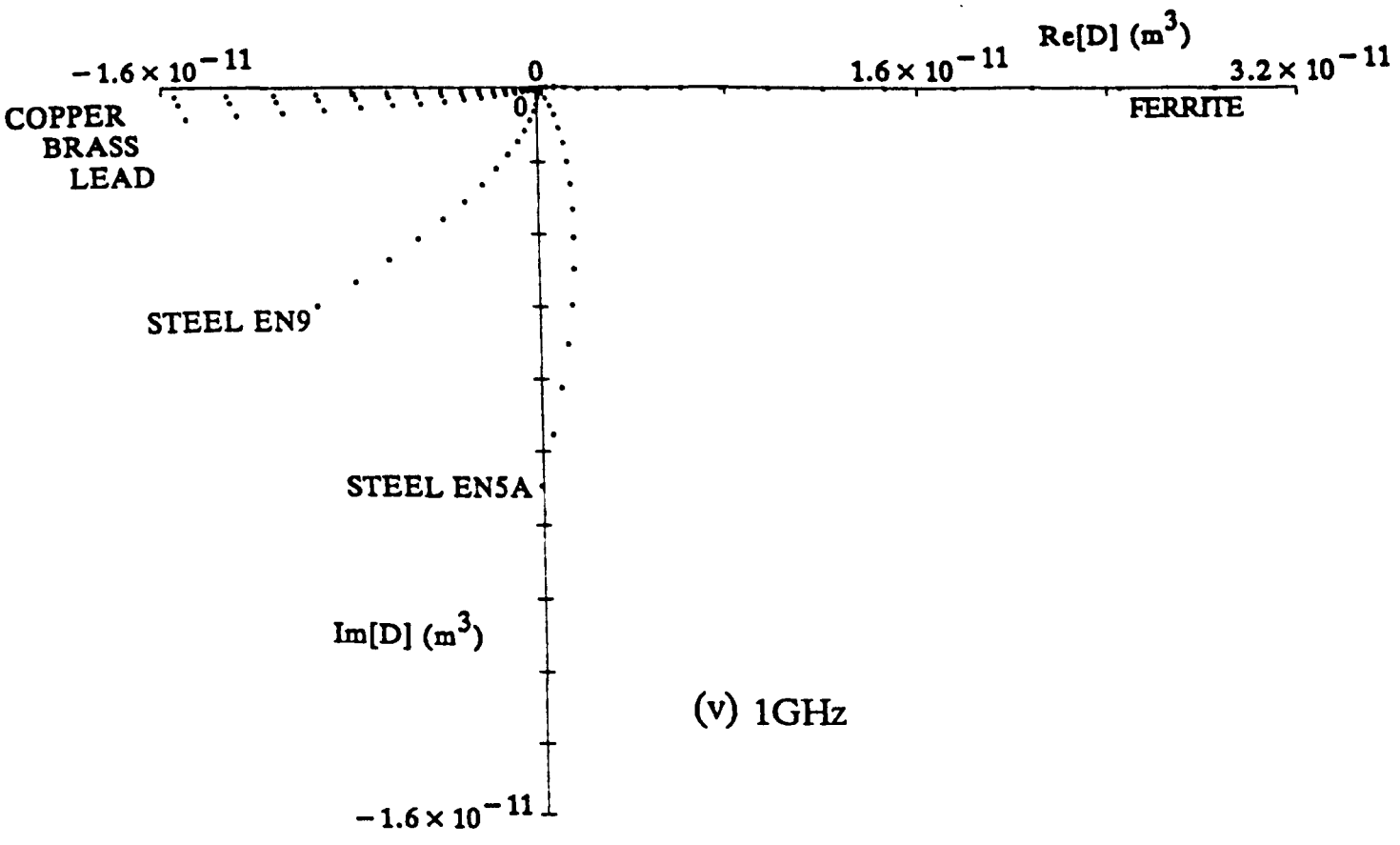


Figure 3.2.4(b) (Continued)

to affect the D_n function much (although he does not define the D_n function explicitly) and indeed his experimental results section describes a short metal cylinder producing the same effect as his theoretical sphere of the same diameter.

In his work on detecting ferrous particles using the Smiths Industries Metal Particle Detector, Lloyd [5] analyses spheroids with an axis in line with the detection coil axis. The effect of the spheroid is described through use of a 'shape factor' which varies according to the eccentricity of the particle. If the particle is not aligned with the detection coil axis then the results are not valid. Moreover, the frequency of the applied field must be such that skin depth limiting does not occur. Consequently, the analysis he presents cannot be used directly here, although his results for spherical samples of high μ_r imply a D_n of 2 which is consistent with this work.

Fromm and Jhen [6] have addressed the problem of varying shapes for levitation melting techniques. They find, experimentally, that the behaviour of their samples, mostly short cylinders, falls within 20% of that expected theoretically from spheres of the same diameter.

The above combination of theoretical and experimental evidence leads us to believe that the shape of particles is not very significant provided that they are roughly spherical. Also, the sphere is the only shape whose scattering field is not orientation dependent. Since we have no control of particle orientation in a lubricant flow, the sphere generates a sort of average effect for a randomly orientated particle of arbitrary shape.

3.3 MAGNETIC FIELDS IN THE INDUCTIVE COIL TRANSDUCER

In §3.1 we outlined reasons for selection of the inductive coil for use as a transducer for lubricant monitoring and we have chosen to work with the single-layer helical coil. The magnetic flux density, **B**, varies depending upon the position within the coil. We now investigate the variation in the 'z-component' of the flux density (see Figure 3.3(a) for the coordinate system used) along the coil axis (i.e. $B_z(0,z)$) and across the coil cross section at $z=0$ (i.e. $B_z(r,0)$). By symmetry, the **B** vector can have only a 'z-component' at these points.

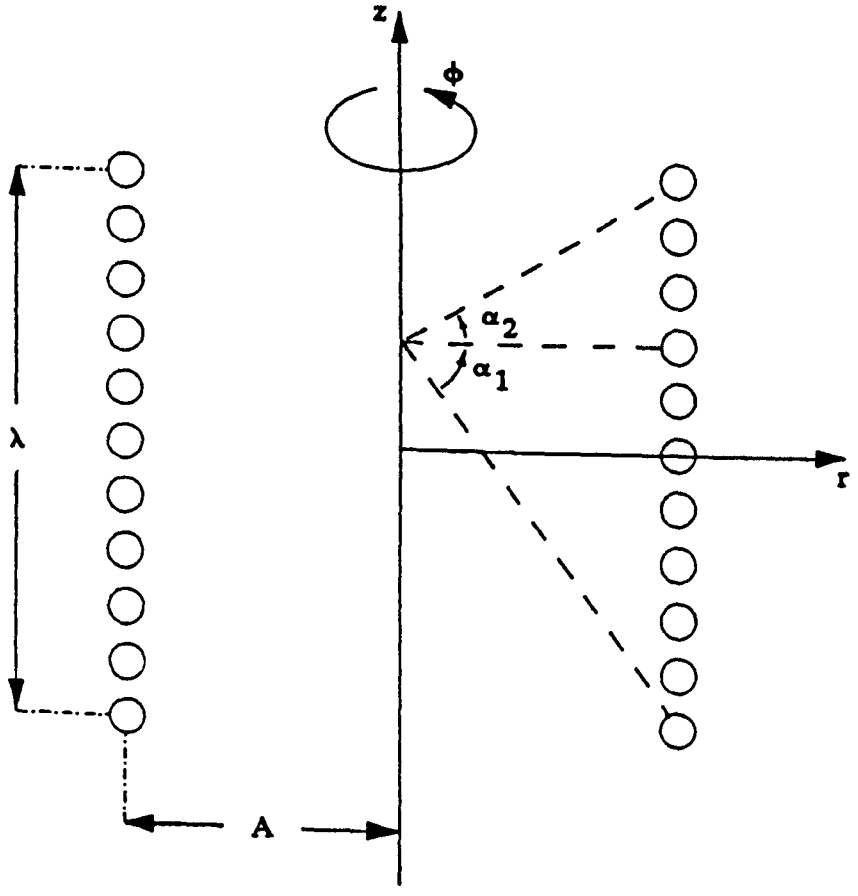


Figure 3.3(a) Schematic diagram of coil cross-section, with coordinates system.

3.3.1 The Flux Density on the z-axis

The variation of flux density along the coil axis is well known [7] and can be approximated by

$$B(0,z) = B_z(0,z)\,a_z = \frac{\mu_o n I_c}{2}(\sin\alpha_2 - \sin\alpha_1)a_z \qquad (1)$$

where α_1 and α_2 are as marked in Figure 3.3(a), n is the coil turns density (units: turns/m) and I_c is the coil current. The assumption made in the approximation is that the coil can be represented by an infinitely thin cylindrical current sheet with a current sheet density of magnitude $|K| = nI_c$ and this holds reasonably true when the number of turns is large and the winding wire diameter is much smaller than coil dimensions.

3.3.2 The Flux Density on the r-axis

A theoretical analysis of radial flux density variation does not appear to exist in the literature and we have investigated the effect here.

Assuming, as per §3.3.1, that the coil can be represented by a cylindrical current sheet we may first obtain the magnetic vector potential A(r,0) at each point to be analysed and by invoking the relationship $B = \nabla \times A$, obtain B.

The magnetic vector potential at the z=0 plane due to the cylindrical current sheet is given by (see Appendix 1)

$$|A(r,0)| = \frac{\mu_o n I_c A}{4\pi} \int_{z=-\lambda/2}^{z=\lambda/2} \int_{\phi=0}^{\phi=2\pi} \frac{\cos\phi}{(z^2+A^2+r^2-2A r\cos\phi)^{1/2}}\, d\phi dz \qquad (1)$$

This equation may be evaluated numerically and we have done this using Simpson's Rule twice with 31 ordinates for each of three different coil geometries. In each case values of $B = B_z(r,0)\,a_z$ have been obtained using the relation $B = \nabla \times A$ within a cylindrical coordinate system, and normalised values given by

$$B_{nz}(r,0) = \frac{B_z(r,0)}{B_\infty} \qquad (2)$$

where $B_\infty = \mu_o n I_c$ denotes the magnitude of the flux density within an infinite coil with the same values of n and I_c and plots of B_{nz} are shown in Figure 3.3.2(a) for three coil shapes. It is clear that the greater the length : radius ratio the more uniform is the coil field and this is well known.

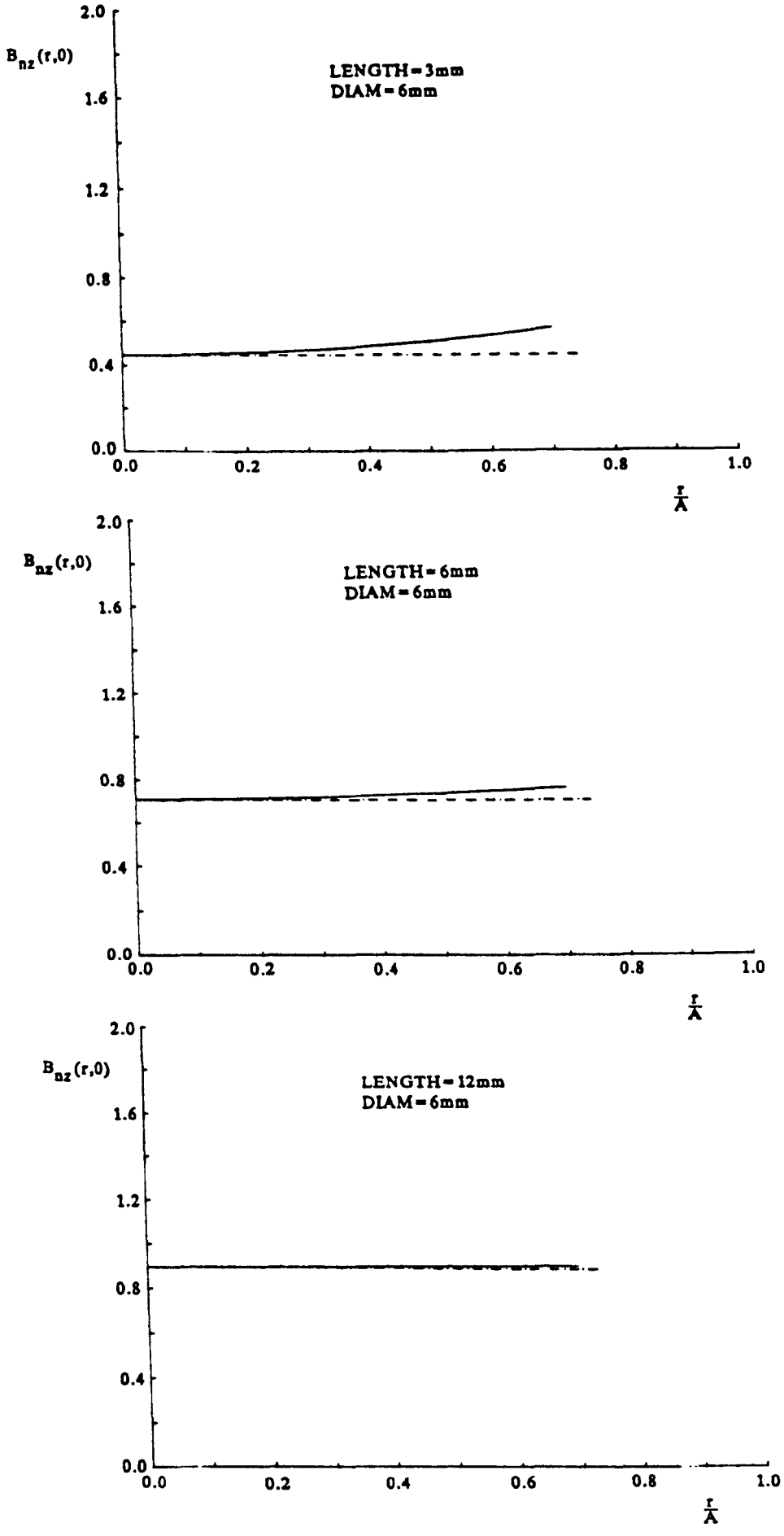


Figure 3.3.2(a) Normalised flux density versus normalised radial distance from coil centre for various coil shapes. Values at normalised radial distances greater than 0.7 are not shown because our model becomes increasingly inaccurate in that region.

3.4 THE CHANGE IN COIL TRANSDUCER IMPEDANCE DUE TO A PARTICLE

Consider a coil transducer fed with an alternating current source, as shown in Figure 3.4(a). The magnitude and phase of the voltage generated across an air-cored coil will represent the coil impedance. As is well known, this impedance may be expressed in terms of two principle components:

(i) A positive reactive component due to coil flux-linkages which are in-phase with the applied current source.

(ii) A positive resistive component due to coil resistance and, with alternating currents, losses due to eddy current phenomena in the windings.

These components are termed $j\omega L$ and R.

As discussed in §3.2.1, we may replace a particle which is subjected to a magnetic field, by an effective magnetic dipole. If such a dipole is placed within the coil of Figure 3.4(a), a voltage change occurs across the coil terminals. Two forms of impedance change take place, their relative contributions depending upon the phase of the effective dipole field.

(i) A change in the reactive component.

(ii) A change in the resistive component.

We label the changes $j\omega\Delta L$ and ΔR respectively and proceed to determine the relationship between them and the coil-particle combination for various particle and coil characteristics. Also, we include a method for obtaining an approximate value of R.

3.4.1 The Particle at the Coil Centre

Consider a particle positioned at the coil centre (i.e. at the point r=0, z=0 in Figure 3.3(a)). We may replace the particle by an equivalent dipole of strength:

$$m = \frac{2\pi B_z D}{\mu_o} \qquad (1)$$

The dipole field produces flux-linkages with the coil windings which generate an impedance change.

The phase of the dipole field relative to that of the original coil field controls the nature of the impedance change: the in-phase component of the dipole field results in a

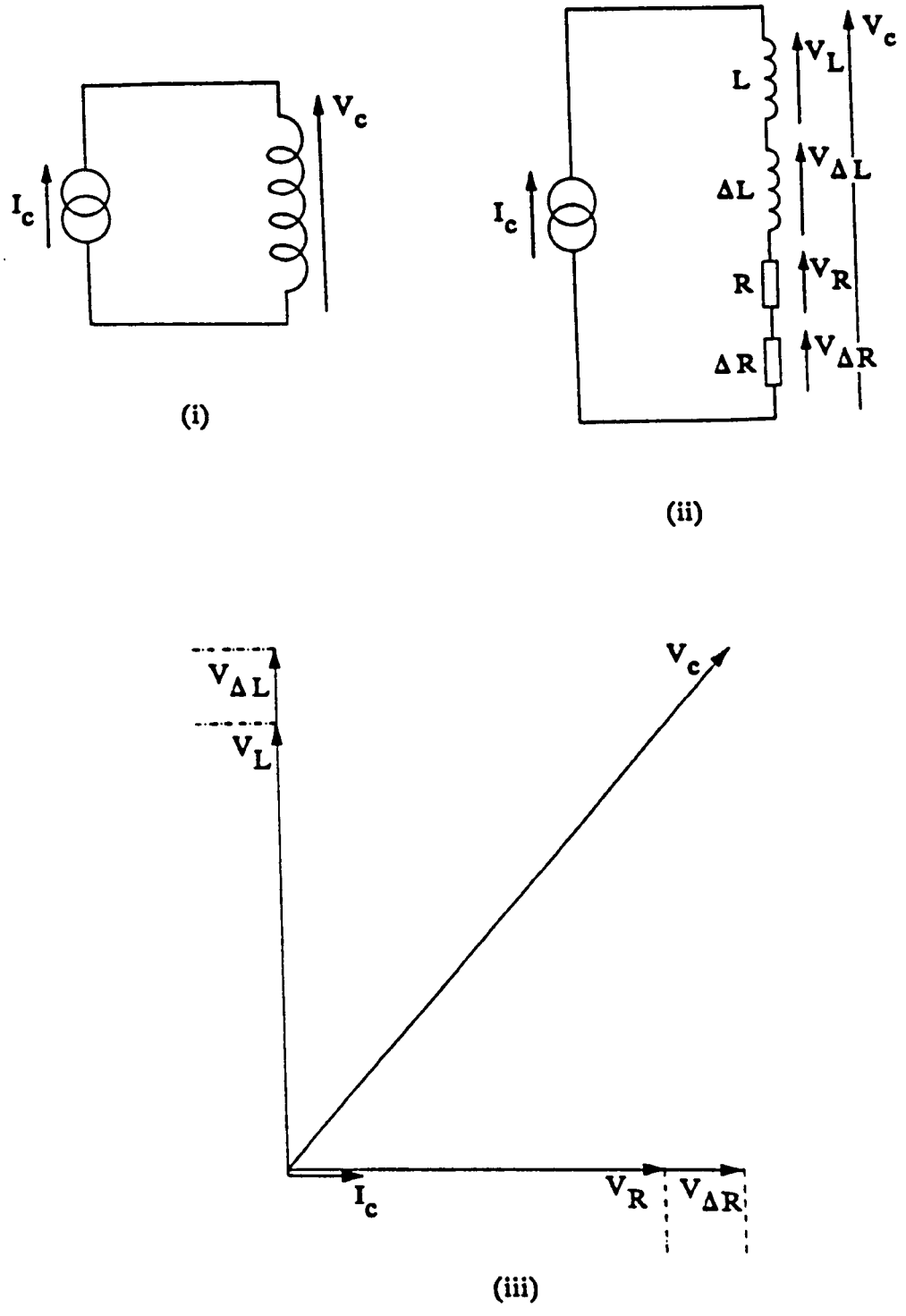


Figure 3.4(a) (i) A schematic diagram of coil transducer fed with an AC current source, (ii) the equivalent circuit and (iii) the associated phasor diagram.

change, ΔL, in inductance whilst the quadrature component results in a change, ΔR, in coil series resistance.

The flux-linkage, Λ_{po}, due to a dipole of strength m at the centre of a coil of length λ and radius A is given by:

$$\Lambda_{po} = n\mu_o m \sin\alpha_o \tag{2}$$

where

$$\alpha_o = \sin^{-1}\left[\frac{(\lambda/2)^2}{(\lambda/2)^2+A^2}\right]^{1/2} \tag{3}$$

Substituting equations 3.4.1(1) and 3.3(2), yields

$$\Lambda_{po} = 2\pi n^2 \mu_o I_c D \sin^2\alpha_o \tag{4}$$

The in-phase component of Λ_{po}, i.e. $Re[\Lambda_{po}]$, adds an inductive component to the coil impedance given by

$$\Delta L = Re[\Lambda_{po}]/I_c = 2\pi n^2 \mu_o Re[D] \sin^2\alpha_o \tag{5}$$

The coil inductance can be approximated by [8]

$$L = K_n\left[\mu_o n^2 \pi A^2 \lambda\right] \tag{6}$$

where K_n is the Nagaoka constant given by

$$K_n = \frac{1}{1+0.9(A/\lambda)-0.02(A/\lambda)^2} \tag{7}$$

Hence

$$\frac{\Delta L}{L} = \left[\frac{2\sin^2\alpha_o}{K_n \lambda A^2}\right] Re[D] \tag{8}$$

We may define the square bracketed term to be a particular case of the function $K_c(r,z)$ for the particle positioned at the point $r=z=0$ and Equation (8) may be rewritten

$$\frac{\Delta L}{L} = K_c(0,0)\, Re[D] \tag{9}$$

(For brevity later, we shall write K_{co} for $K_c(0,0)$.)

Rearranging Equation (8) yields

$$Re[D] = \left[\frac{K_n \lambda A^2}{2\sin^2\alpha_o}\right] \frac{\Delta L}{L} \tag{10}$$

Similarly, $\mathrm{Im}[\Lambda_{po}]$ generates a voltage, V_R, which using Faraday's law is

$$V_R = -j\omega\ \mathrm{Im}[\Lambda_{po}] = -j\omega\ 2\pi n^2\mu_o\ I_c\ \mathrm{Im}[D]\ \sin^2\alpha_o \tag{11}$$

or, equivalently

$$\Delta R = -\frac{|V_R|}{I_c} = 2\pi\omega n^2\mu_o\ \mathrm{Im}[D]\ \sin^2\alpha_o \tag{12}$$

Hence

$$\mathrm{Im}[D] = -\left[\frac{K_n\lambda A^2}{2\sin^2\alpha_o}\right]\frac{\Delta R}{\omega L} \tag{13}$$

or

$$\frac{\Delta R}{\omega L} = -K_c(0,0)\ \mathrm{Im}[D] \tag{14}$$

We have thus obtained, for a spherical metal particle at the coil centre, relationships between the observable impedance changes and the complex D function components.

3.4.2 The Particle on the z-axis

Consider a particle on the coil axis (i.e. the z-axis of Figure 3.3(a)). The flux linkage, $\Lambda_p(0,z)$, due to a particle on the axis of a coil of length λ and radius A is given by

$$\Lambda_p(0,z) = \frac{n\mu_o m}{2}(\sin\alpha_2 - \sin\alpha_1) \tag{1}$$

where α_1 and α_2 are positional variables as marked in Figure 3.3(a).

Substituting equations 3.4.1(1) and 3.3.1(1) into the above equation yields

$$\Lambda_p(0,z) = \frac{\mu_o n^2 2\pi I_c D}{4}(\sin\alpha_2 - \sin\alpha_1)^2 \tag{2}$$

Hence, using equation 3.4.1(4)

$$\frac{\Lambda_p(0,z)}{\Lambda_p(0,0)} = \frac{(\sin\alpha_2 - \sin\alpha_1)^2}{4\sin^2\alpha_o} \tag{3}$$

where α_o is a function of coil shape given by equation 3.4.1(3).

Thus

$$\Delta L = \frac{\mathrm{Re}[\Lambda_{po}]}{I_c} = \frac{\pi n^2\mu_o \mathrm{Re}[D](\sin\alpha_2 - \sin\alpha_1)^2}{2} \tag{4}$$

and

$$\mathrm{Re}[D] = \left[\frac{2K_n\lambda A^2}{(\sin\alpha_2 - \sin\alpha_2)^2}\right]\frac{\Delta L}{L} \tag{5}$$

or

$$\frac{\Delta L}{L} = K_c(0,z)\ \mathrm{Re}[D] \tag{6}$$

Similarly

$$\mathrm{Im}[D] = -\left[\frac{2K_n\lambda A^2}{(\sin\alpha_2 - \sin\alpha_1)^2}\right]\frac{\Delta R}{\omega L} \tag{7}$$

or

$$\frac{\Delta R}{\omega L} = -K_c(0,z)\ \mathrm{Im}[D] \tag{8}$$

Equations (5) and (7) naturally reduce to equations 3.4.1(10) and 3.4.1(13) for a particle at the coil centre.

The ratio of the flux linkages due to a particle on the z-axis of a coil of arbitrary length and radius to those due to the same particle at the centre of an infinite coil with the same turns density and current is given by

$$\Lambda_{np}(0,z) = \frac{\Lambda_p(0,z)}{\Lambda_{\infty p}} = \frac{(\sin\alpha_2 - \sin\alpha_1)^2}{4} \tag{9}$$

and this is plotted in Figure 3.4.2(a) for three shapes of coil.

3.4.3 The Particle on the r-axis

Consider a particle on the r-axis of Figure 3.3(a). The flux linkage, $\Lambda_p(r,0)$, due to the equivalent dipole of strength m on the r-axis within a coil of length λ and radius A is given by (see Appendix 2)

$$\Lambda_p(r,0) = \frac{\mu_o mn}{4\pi}\int_{z=-\lambda}^{\lambda}\int_{\phi=0}^{2\pi}\frac{(A\cos\phi - r)^2 + (A\sin\phi)^2 + A^2 - r^2}{2[(A\cos\phi - r)^2 + (A\sin\phi)^2 + z^2]^{3/2}}\,d\phi\,dz \tag{1}$$

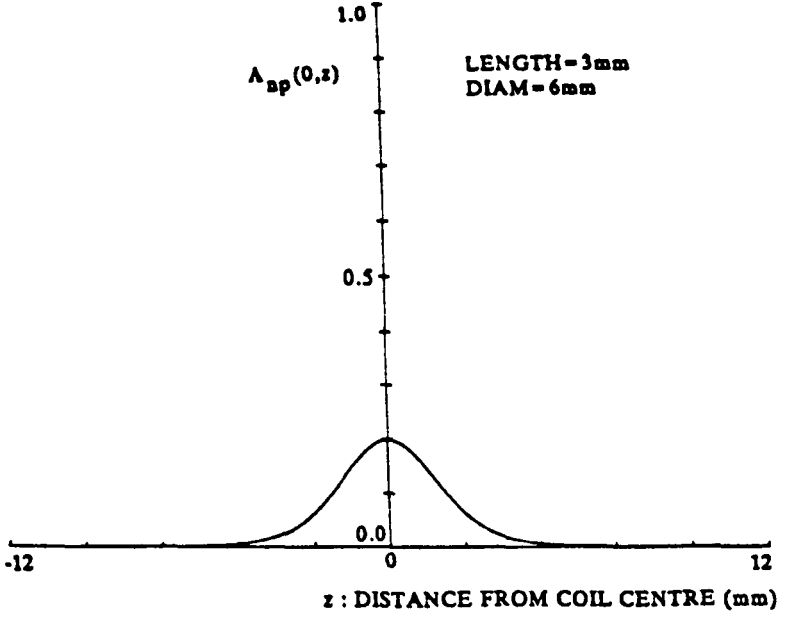
Also, from §3.3.2

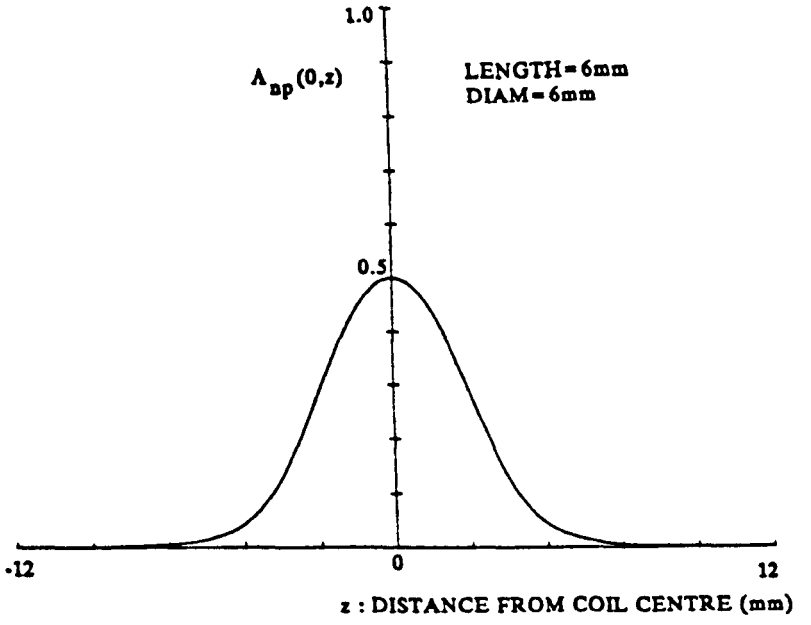
$$m = \frac{2\pi B_z D}{\mu_o} \tag{2}$$

$$B_z(r,0) = |\nabla \times A(r,0)| \tag{3}$$

and A(r,0) is given by equation 3.3.2(1).

Equation (1) has been evaluated using Simpson's Rule with 31 ordinates and combining the above equations yields the change in flux linkages due to a particle on





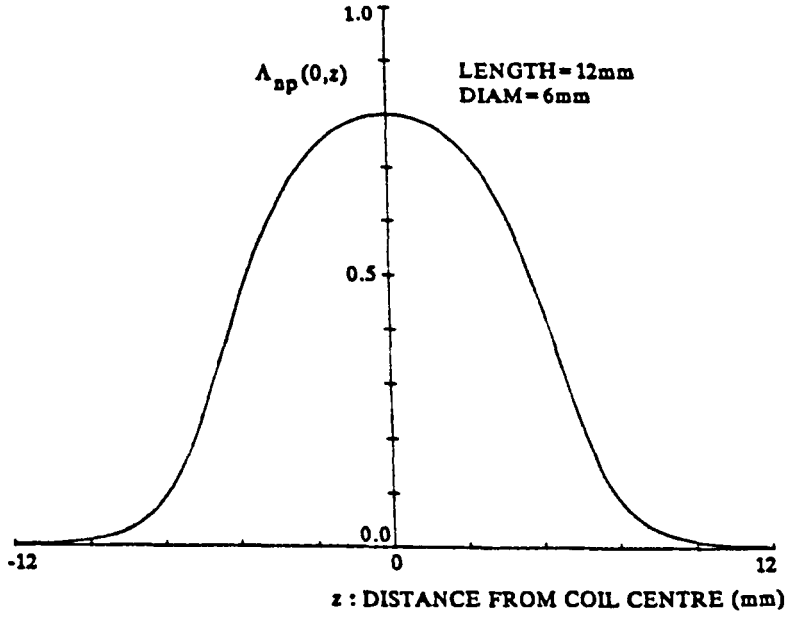


Figure 3.4.2(a) Normalised flux-linkage for particle on the coil axis (i.e. r=0) for various coil shapes.

the r-axis. As per §3.4.2 we have normalised these values to the flux linkage change due to the same particle at the centre of an infinite coil with the same turn density and current to yield

$$\Lambda_{np}(r,0) = \frac{\Lambda_p(r,0)}{\Lambda_{\infty p}} \qquad (4)$$

Plots of this function for each of three coil shapes are shown in Figure 3.4.3(a).

3.4.4 Coil Transducer Series Resistance

The value of coil series resistance, R, has not appeared in the equations describing the effects of a metal particle in the coil transducer but is nevertheless essential to the analyses of system characteristics which are included in subsequent chapters.

An analytic method for determining the series resistance of an inductance coil of known geometry and material is not evident in the literature and Welsby [8] refers to the empirical work of Medhurst [9]. This applies to the situation where a single-layer coil is constructed with closely-packed windings (i.e. where the ratio $d/c > 0.5$, the value d being the winding wire diameter and c the spacing between the axes of adjacent windings) and operated at a high frequency (i.e. where the ratio $d/\delta > 5$, the parameter δ being the skin-depth in the winding material).

The value of R is then given approximately by

$$R = \left(\frac{1}{2}\right).R_{dc}.\left(\frac{d}{\delta}\right).\Phi \qquad (1)$$

where R_{dc} is the DC resistance of the coil and Φ represents an empirical function plotted by Medhurst and reproduced here as Figure 3.4.4(a). The function depends on both the coil length to diameter ratio and the d/c ratio. To minimise series resistance we would require to minimise the d/c ratio but this would, for a given desired inductance, require an increase in coil length.

3.5 SUMMARY AND CONCLUSIONS FOR CHAPTER THREE

This chapter is summarised and concluded as follows:

- The inductive coil transducer is the most appropriate for monitoring small metal particles in the approximate size range 100-500μm.
- A spherical metal particle immersed in a uniform magnetic field of arbitrary frequency within a vacuum may be ascribed a complex value D (units m^3) which is a function of magnetic flux density, operating frequency, particle diameter and

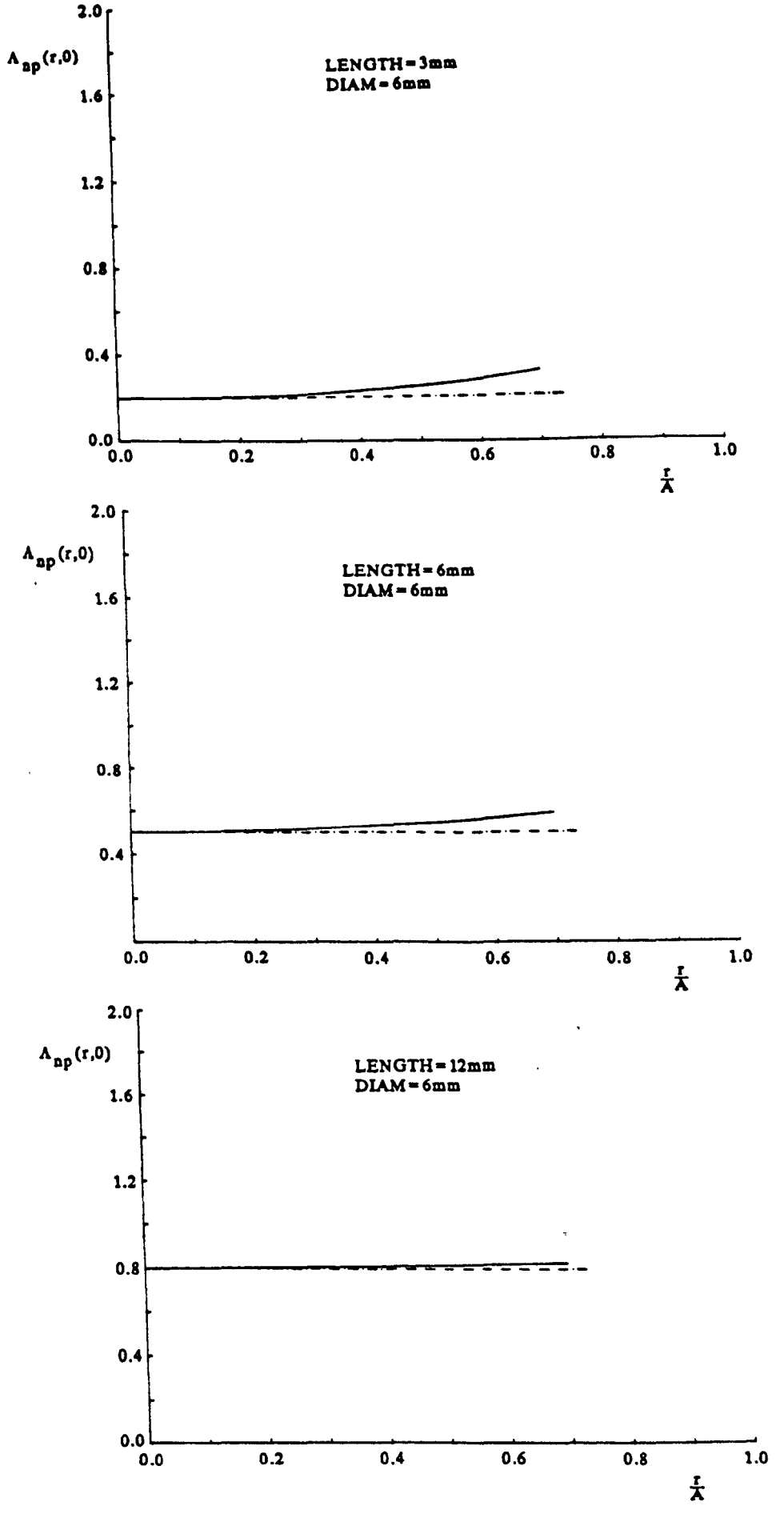


Figure 3.4.3(a) Normalised flux-linkage on r-axis on coil central transverse plane (i.e. z=0) for various coil shapes. Values at normalised radial distances greater than 0.7 are not shown because our model becomes increasingly inaccurate in that region.

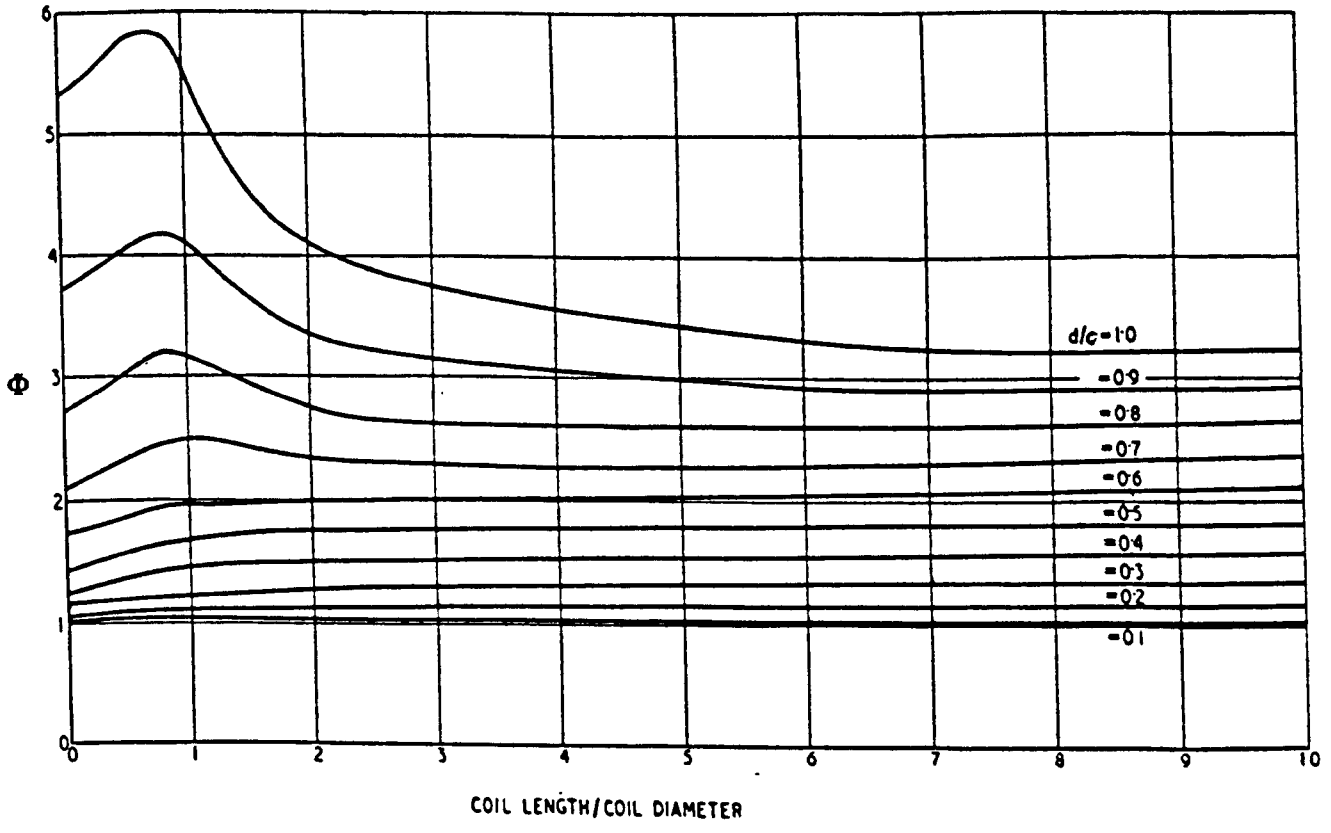


Figure 3.4.4(a) Plot of function Φ for various coil shapes and winding spacings [9].

the values of the particle material resistivity and magnetic permeability.

- Operation at ~10MHz was selected for the present system as this gives a good spread of D function loci for materials of interest.
- Such a particle exhibits the properties of a magnetic dipole of strength $m=\frac{2\pi B_z D}{\mu_o}$ at distances from the particle much greater than the particle diameter. Moreover, the analysis for a spherical particle can be extended, with reasonable accuracy, to any roughly spherical particle.
- When such a particle is passed through an energised inductance coil the effective dipole interacts with the coil windings to produce perturbation pulses in the coil series resistance and inductance given by

$$\frac{\Delta R}{\omega L} = -K_c(r,z)\ \mathrm{Im}[D]$$

and

$$\frac{\Delta L}{L} = K_c(r,z)\ \mathrm{Re}[D]$$

where K_c is a function of position within the coil as well as coil size and shape. A value $K_c(0,0)$ is used to compute the effect of the particle at the coil centre and is a function of coil geometry.

- The magnitude of the perturbations due to a particle passing through the coil is not strongly dependent on the radial distance of the particle from the coil centre provided that the particle is restricted to within, say, half a coil radius from the coil axis.
- The values of coil series resistance and inductance are obtained from simple formulae.

CHAPTER THREE: REFERENCES

1. Lim D H: 'A capacitive method for lubricant monitoring', BSc(Hons) Project Report HSP387, University of Edinburgh, May 1984. (Unpublished).

2. Smythe W R: 'Static and Dynamic Electricity', McGraw-Hill, New York, 2nd ed (1950) Chapter 11.

3. Wejgaard W and Tomar V S: 'Contactless measurement of electrical resistivity by eddy current methods: discussion and a new technique', J Phys E: Sci Instrum, 7 (1974) 395-9.

4. Zimmerman J E: 'Measurement of electrical resistivity of bulk metals', Rev Sci Instrum, **32** (1961) 402-5.

5. Lloyd O, 'Wear debris in lubrication systems; an evaluation of a Smiths Industries in-line chip detector', CEGB Marchwood Engineering Laboratories Report R/M/N1094, March 1980.

6. Fromm E and Jhen H: 'Electromagnetic forces and power absorption levitation melting', Brit J Appl Phys, **16** (1965) 653-63.

7. Rorschach H E and Herlin M A: 'An induction method for measuring electrical resistivity', Mass Inst Technol Research Lab Electronics Tech Rept No 125 (12 October 1952).

8. Welsby V G, The Theory and Design of Inductance Coils, 2nd ed, Macdonald, London, (1960).

9. Medhurst R G: 'H.F. resistance and self-capacitance of single-layer solenoids', Wireless Engineer, **24** (1947) 80-92.

CHAPTER FOUR

DRIVE CIRCUITRY FOR INDUCTIVE COIL TRANSDUCERS

Having decided on the use of an inductive transducer, we now consider appropriate drive circuitry. In this chapter we discuss various possible drive circuitry techniques and justify the selection of the marginal oscillator for this work. The operation of marginal oscillators is described in general and with reference to specific circuits used.

4.1 THE MARGINAL OSCILLATOR AND OTHER TYPES OF TRANSDUCER DRIVE CIRCUITRY

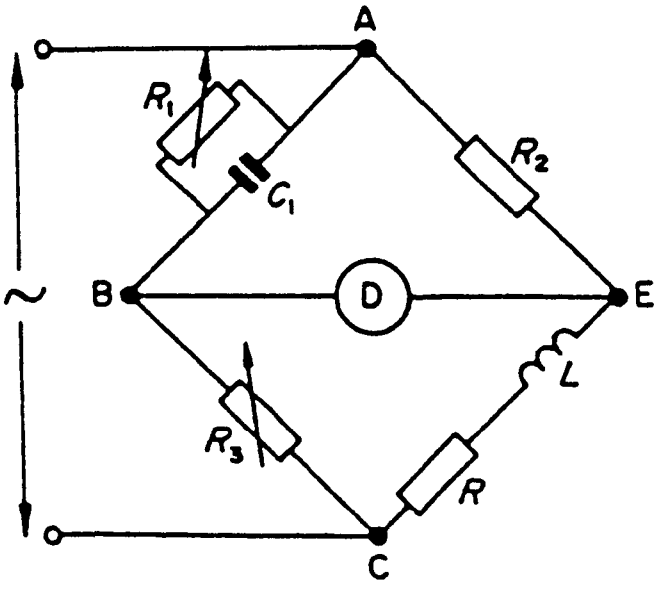
The purpose of the transducer drive circuitry is to obtain a signal proportional to the change in coil transducer resistance and inductance due to a metal wear-debris particle passing through the coil. The most common technique used for detecting such impedance changes is through use of some form of bridge circuit, although decay and oscillator techniques are also used. We shall consider each in turn.

4.1.1 Bridge Circuit Methods

The types of bridge circuit most appropriate for detecting the resistance and inductance of an inductive coil are the Maxwell and Owen bridges [1], as shown in Figures 4.1.1(a) and (b). In each circuit the transducer forms an unknown arm. For the Owen bridge, determining the values of resistor R_3 and capacitor C_3 which are required to obtain a zero reading on the detector, and noting fixed component values C_1 and R_2, we may obtain R and L. For the Maxwell bridge resistors R_1 and R_3 must be adjusted to obtain balance and taking the values C_1 and R_2 into account yields R and L.

Prior to use, we could balance the bridge by adjusting the appropriate variable components. Perturbations in the coil impedance due to a metal particle would then cause a deviation of the detector voltage from 0V. The pulse height would depend on a *combination* of the changes in R and L and we would obtain only one piece of information on the particle as we cannot decouple the two effects.

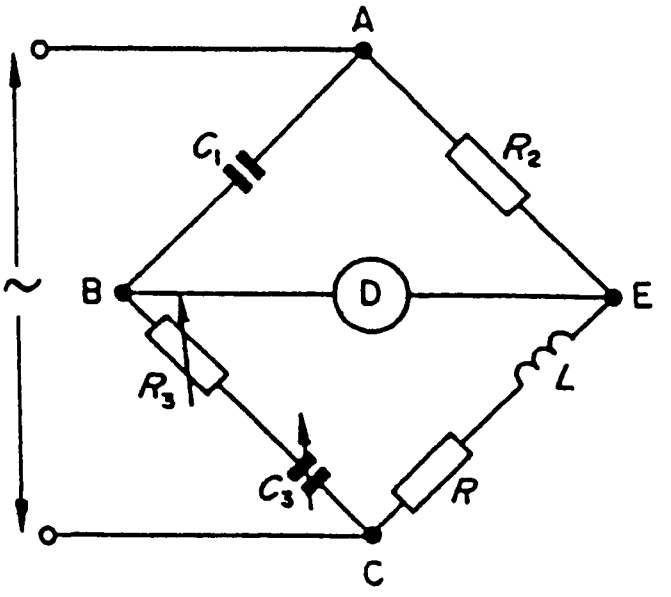
Another type of bridge circuit, the transformer bridge [1], could be used but again the response from the detector would depend on a combination of the resistive and inductive perturbations.



$$L = R_2 R_3 C_1$$

$$R = \frac{R_2 R_3}{R_1}$$

Figure 4.1.1(a) Maxwell Bridge [1]



$$L = R_2 R_3 C_1$$

$$R = \frac{R_2 C_1}{C_3}$$

Figure 4.1.1(b) Owen Bridge [1]

4.1.2 Decay Methods

An alternative approach to finding impedance changes due to interaction with a metal particle could use decay methods. These depend on obtaining the time constant of the coil, T=L/R, by monitoring the the transient current waveform due to an applied step or impulse voltage waveform. Clearly, a measure of the time constant could be obtained, but as with the bridge circuit approach we would obtain a response dependent on a *combination* of changes in R and L.

4.1.3 The Marginal Oscillator Method

A third option is the use of the marginal oscillator (see Figure 4.1.3(a)) which enables us to measure, to a good approximation, the changes in R and L *independently*. Changes in L cause changes in oscillator frequency, whilst changes in R alter the amplitude of the oscillations. By frequency and amplitude demodulating the signal monitored at a convenient point in the oscillator loop, we can obtain two pieces of information on each particle which passes through the coil. This is the principle reason for selection of the marginal oscillator in preference to bridge or decay methods.

4.2 MARGINAL OSCILLATOR SIGNAL AND NOISE CHARACTERISTICS

We define a marginal oscillator to be a resonant oscillator whose operating level is controlled by the nonlinearity of the transfer characteristic of the oscillator gain element. Referring to the model shown in Figure 4.1.3(a), the gain of the transconductance amplifier is nonlinear and the initial signal build-up is stabilised by the reduction in gain as the signal level increases. The peak steady-state signal level, V_o, is obtained when the amplifier transconductance equals the conductance, G, of the tuned circuit.

Marginal oscillators respond to small changes in tuned circuit conductance, ΔG, with a change in oscillation amplitude, ΔV_o (which we define as the change in the *peak* value of the fundamental voltage signal across the tuned circuit) whilst changes in tuned circuit reactance cause a shift in the frequency of oscillation. Spurious noise adds to these desired signal effects and limits the sensitivity of the system. For this section we have developed general equations which allow us to relate the effects of specific particles to system frequency and amplitude response and also to relate spurious noise to output signal-to-noise ratios.

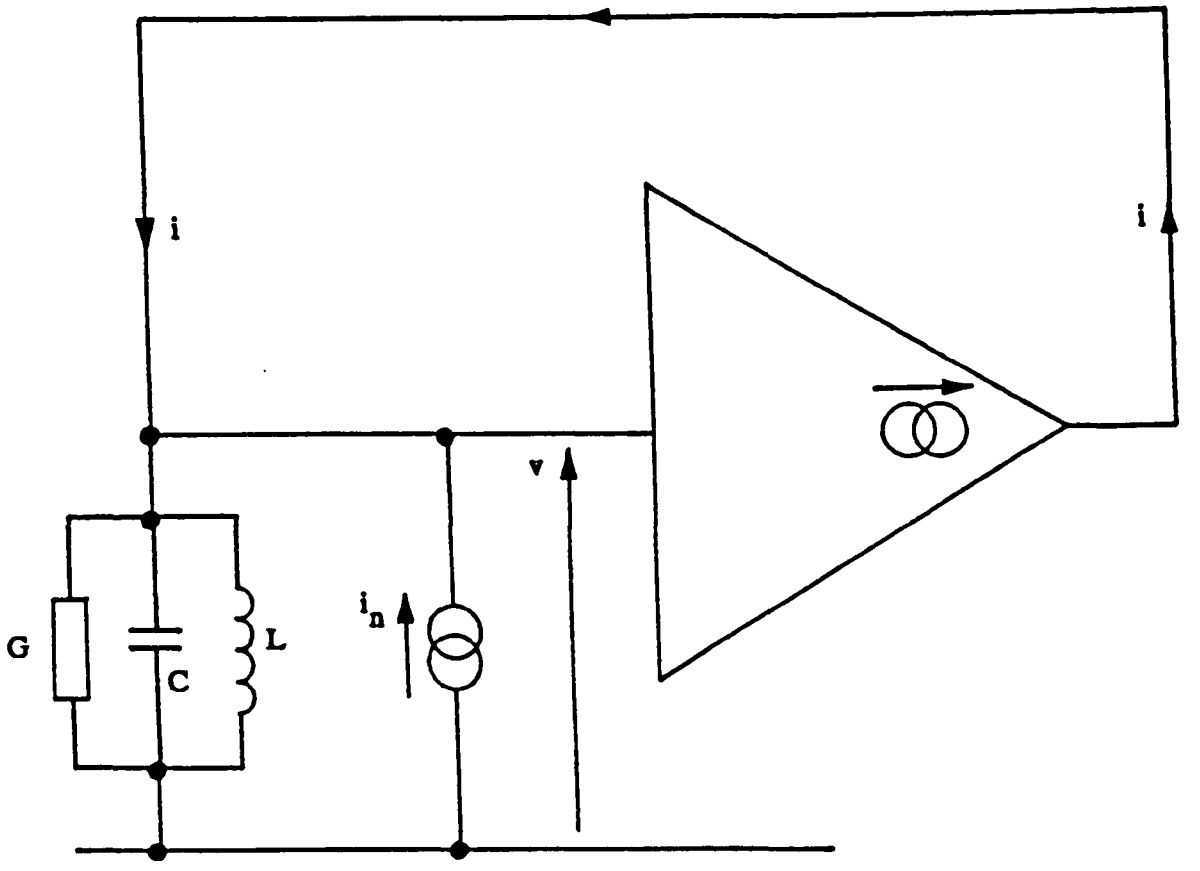


Figure 4.1.3(a) A schematic circuit diagram of a marginal oscillator with the nonlinear amplifier configured as a transconductance amplifier. Current variable, i, and voltage variable, v, represent *instantaneous* values of the fundamental components of the waveforms at the points shown. The noise current source, i_n, represents the equivalent noise effects due to the tank circuit and amplifier.

4.2.1 Amplitude Signal Perturbation

Amplitude signal perturbation effects are due mainly to the change in coil series resistance, R, which, for a coil in a resonant tuned circuit, is equivalent to a change in the effective parallel tuned circuit conductance, G. The values of R and G are related by [2]

$$G = (C/L)\,R \tag{1}$$

and differential changes by

$$dG = (C/L)\,dR \tag{2}$$

for a tuned circuit operating at the resonant frequency (see Figure 4.2.1(a)).

From geometric considerations, the differential change in oscillation amplitude, dV_o, may be related to the change in tuned circuit conductance, dG, by the relationship

$$dV = -\left(\frac{V_o}{G-T(V_o)}\right) dG \tag{3}$$

where $T(V_o)$ is the gradient of the I_o-V_o characteristic of the nonlinear amplifier at the working point (see Figure 4.2.1(b)). This equation may be rewritten as

$$\left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right) = -\left(\frac{G}{G-T(V_o)}\right) \tag{4}$$

Robinson [3] uses the factor $(G/V_o)(dV_o/dG)$ in developing his oscillator amplitude theory and our approach is based on his. The factor exhibits a 3dB low-pass characteristic with a cut-off frequency given by [3]

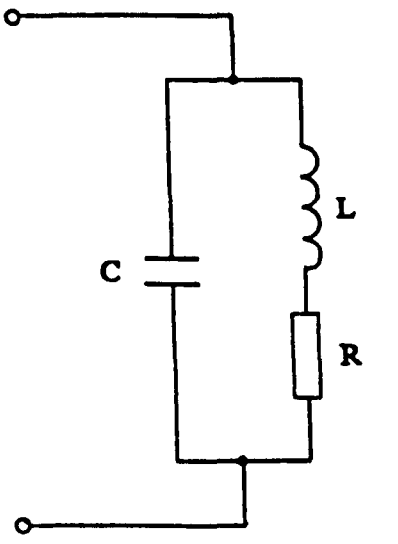
$$f_{co} = \frac{f_o}{2Q(G/V_o)(dV_o/dG)} \tag{5}$$

Provided that we design circuitry such that perturbations occur with spectral components well below this cut-off frequency we can assume Equation (4) to be frequency independent and the following analysis is done on this basis.

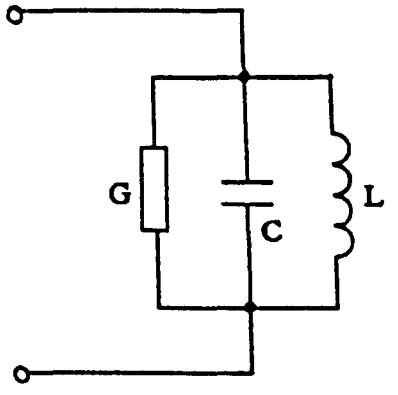
From Equation (4) we note that the most sensitive arrangement is one in which the value of $T(V_o)$ is close to that of G and Figure 4.2.1(c)(i) shows van der Pol oscillator characteristics which illustrate the effect. The van der Pol oscillator is usually specified as an oscillator controlled by a characteristic of the form

$$i = H_1v + H_2v^2 + H_3v^3 \tag{6}$$

where v and i are the *instantaneous* values of the transconductance feedback amplifier



(i)



(ii)

Figure 4.2.1(a) Equivalent tank circuit models for a circuit operating at the resonant frequency: (i) the series model and (ii) the parallel model.

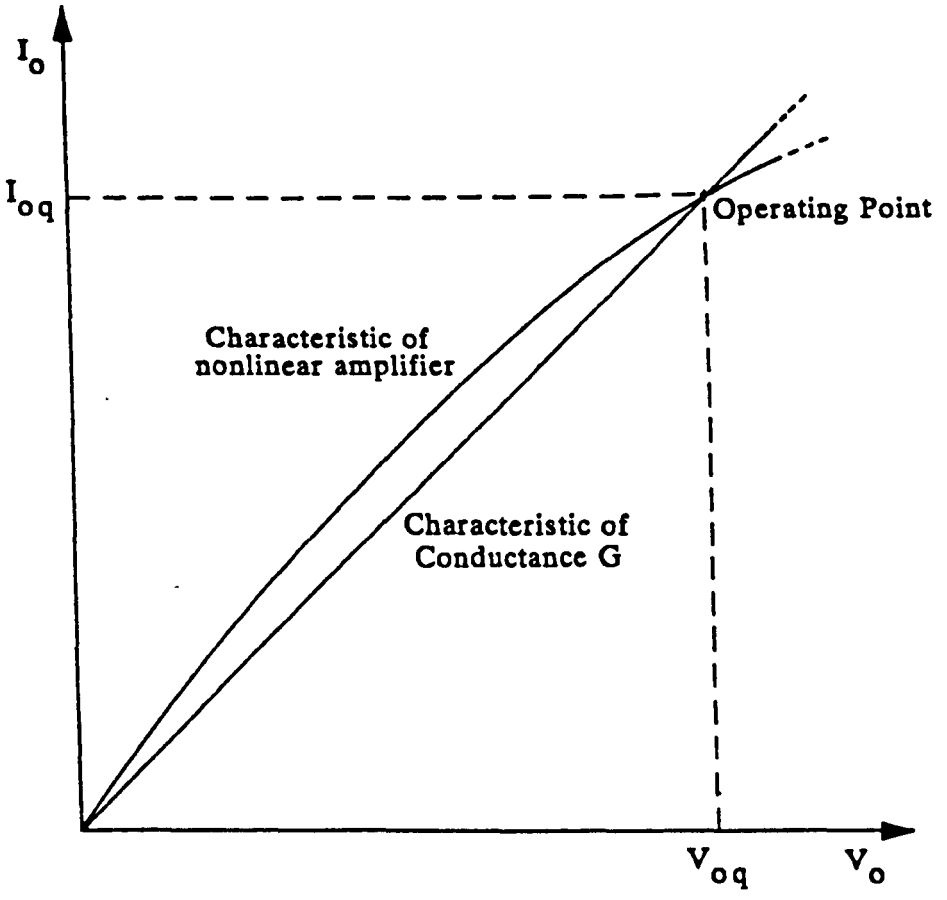
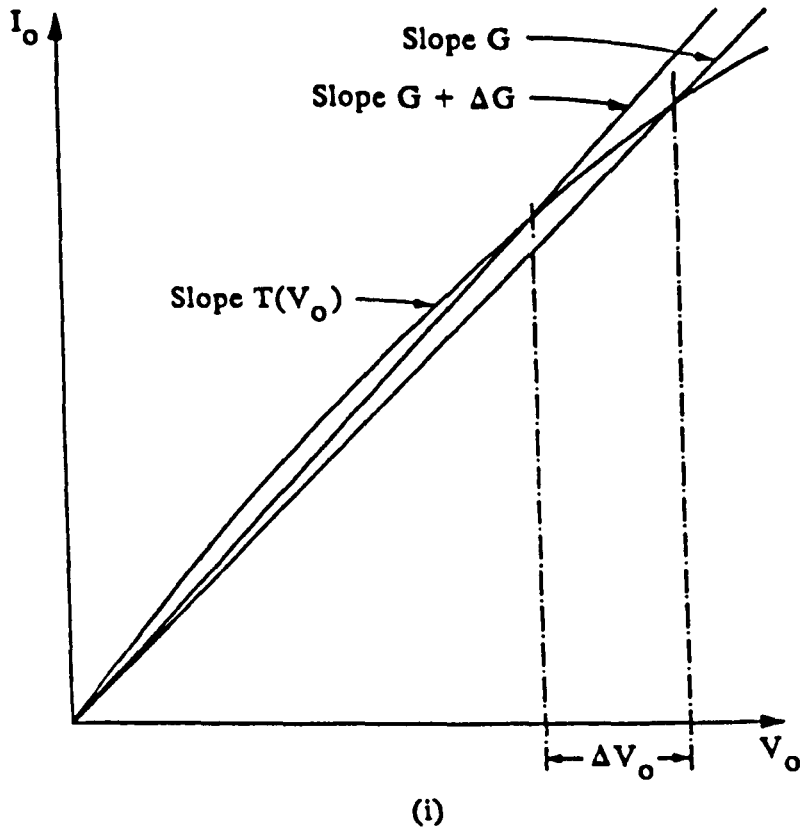
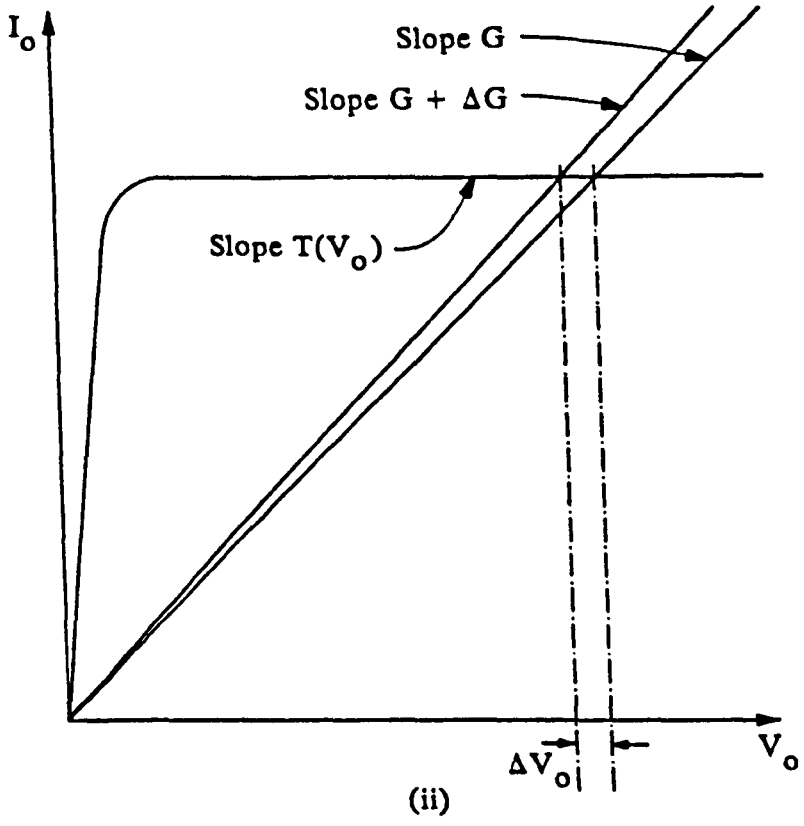


Figure 4.2.1(b) The I_o-V_o characteristics for a marginal oscillator. The intersection of the effective tank conductance characteristic with the nonlinear amplifier characteristic gives the operating point (V_{oq},I_{oq}).



(i)



(ii)

Figure 4.2.1(c) The I_o-V_o characteristics of (i) the van der Pol oscillator and (ii) the Robinson Oscillator with the effect of a change, ΔG, in the tank conductance, G, and the corresponding change, ΔV_o, in the peak osicllator voltage signal, V_o shown in each case.

input voltage and output current. The peak fundamental voltage signal, V_o is then given by [4]

$$V_o = \left(\frac{4}{3}\,\frac{G - H_1}{H_3}\right)^{1/2} \tag{7}$$

and

$$I_o = H_1 V_o + (3/4) H_3 V_o^3 \tag{8}$$

Hence

$$T(V_o) = \frac{dI_o}{dV_o} = H_1 + (9/4) H_3 V_o^2 \tag{9}$$

Thus,

$$\left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right) = -\left(\frac{1}{2}\right)\left(\frac{G}{H_1 - G}\right) \tag{10}$$

By contrast, Figure 4.2.1(c)(ii) shows Robinson oscillator characteristics which exhibit a much lower sensitivity to conductance perturbations. Noting that $T(V_o)=0$ for a Robinson oscillator and using Equation (4) we obtain

$$\left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right) = -1 \tag{11}$$

The van der Pol oscillator is clearly more sensitive to amplitude perturbations than the Robinson oscillator and, indeed, can be made extremely sensitive simply by operating such that the factor H_3, the dominant nonlinearity in the model, is reduced to very low levels. Maximising the amplitude sensitivity does, however, present noise and drift problems and we shall discuss these effects in due course.

4.2.2 Amplitude Noise Perturbation

The noise in a marginal oscillator may be modelled as an additional current source as shown in Figure 4.1.3(a). Such noise current adds vectorially to produce a net current fed into the tuned circuit. Making the usual assumption that the noise is of random phase, the effect of the noise current will be such that equal perturbations of both amplitude and phase will result and we may, for the purpose of analysis, treat the effects independently.

The amplitude noise is due to the in-phase component of the noise current source, and the power spectral density function of the resulting voltage amplitude noise

is given by [3]

$$S_a = \frac{S_i(\omega_o+\omega_m) + S_i(\omega_o-\omega_m)}{G^2} \left(\frac{G}{V_o}.\frac{dV_o}{dG}\right)^2 \quad (1)$$

where $S_i(\omega_o \pm \omega_m)$ are the current power spectral densities of the in-phase component of the noise source (units A^2/Hz). We may integrate S_a over the bandwidth of interest and square-root the result to obtain the overall rms noise voltage output for the oscillator amplitude:

$$V_{n\,rms} = \frac{1}{G}\left(\frac{G}{V_o}.\frac{V_o}{dG}\right)\left[\int_{\omega_m=0}^{\omega_b} S_i(\omega_o+\omega_m) + S_i(\omega_o-\omega_m)\, d\omega_m\right]^{1/2} \quad (2)$$

which may be rewritten as

$$V_{n\,rms} = \frac{1}{G}\left(\frac{G}{V_o}.\frac{dV_o}{dG}\right)\frac{I_{n\,rms}}{2^{1/2}} \quad (3)$$

where $\frac{I_{n\,rms}}{2^{1/2}}$, the rms value of the noise current source of Figure 4.1.3(a), is given by the square-root term in Equation (2).

4.2.3 Frequency Signal Perturbation

Perturbations in the frequency of oscillation are caused by changes in tuned circuit reactance and the interlinking relationship is easily established for small changes. For a marginal oscillator with a high Q tuned circuit the frequency of oscillation is given by

$$\omega_o = \left(\frac{1}{LC}\right)^{1/2} \quad (1)$$

and for the case where C is held constant whilst L is perturbed:

$$\frac{d\omega_o}{\omega_o} = -\frac{1}{2}\frac{dL}{L} \quad (2)$$

As usual, we may replace the differentials by perturbations provided that these are small. In addition, we note that such perturbations are subject to a low-pass filtering effect, the 3dB cut-off frequency being given by

$$f_{co} = \frac{f_o}{2Q} \quad (3)$$

The above equations are independent of the nature of amplitude limiting mechanism of the oscillator.

4.2.4 Frequency Noise Perturbation

From Robinson [3] (or, less directly, from Leeson [5]), the power spectral density of $\dot{\phi}$, the time differential of the phase, is given by

$$S_{\dot{\phi}} = \frac{1}{2}\left(\frac{1}{2CV_o}\right)^2 \left[\int_{\omega_m=0}^{\omega_b} S_i(\omega_o+\omega_m) + S_i(\omega_o-\omega_m)\, d\omega_m\right] \qquad (1)$$

which can be rewritten as

$$S_{\dot{\phi}} = \frac{1}{2}\left(\frac{1}{2CV_o}\right)^2 I^2_{n\,rms} \qquad (2)$$

where $I_{n\,rms}$ is the total low frequency (i.e. $< \frac{\omega_o}{2Q}$) rms quadrature noise current, and this is equal in magnitude to the rms in-phase noise current which generates amplitude noise. Hence, the rms value of the frequency deviation is given by

$$f_{n\,rms} = \frac{1}{4\pi}\left(\frac{1}{2^{\frac{1}{2}}\, C\, V_o}\right) I_{n\,rms} \qquad (3)$$

4.3 MARGINAL OSCILLATOR CIRCUITS IN THE LITERATURE

Several marginal oscillator circuits have been presented in the literature and have used a variety of circuit configurations. The best method of classification is probably on the basis of the component type used in the active stages; i.e. the thermionic valve (or tube), the Metal Oxide Semiconductor Field Effect Transistor (MOSFET), the Junction Field Effect Transistor (JFET) and the integrated amplifier.

Most of the discrete component marginal oscillator circuits of recent years are direct developments of the valve based circuits of the 1950's and 60's. The first major application of marginal oscillators, presented by Pound & Knight in 1950 [6], was for measuring Nuclear Magnetic Resonance absorption effects. Their basic configuration, which uses a dual-valve cathode-follower / common-grid cascade as the nonlinear amplifier, has changed little. Indeed, by the early 1960's the circuit had been only slightly improved [7,8,9] and examples of resulting circuits are shown in Figures 4.3(a)-(c). Such circuits were usually designed for use as van der Pol oscillators

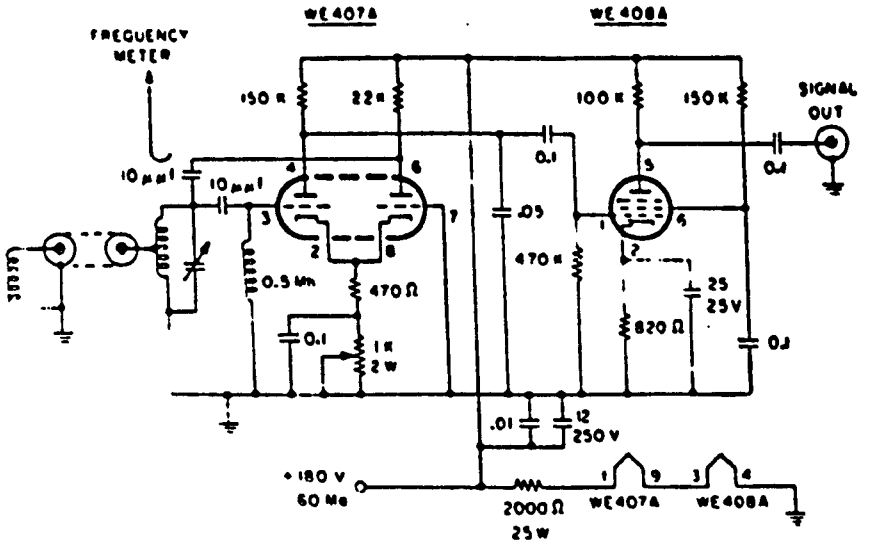


Figure 4.3(a) The marginal oscillator circuit due to Buss & Bogart (1960) [7].

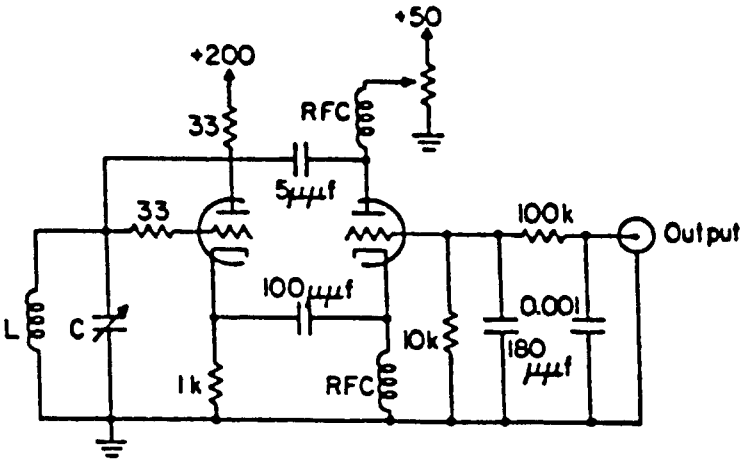


Figure 4.3(b) The marginal oscillator circuit due to Knight (1960) [8].

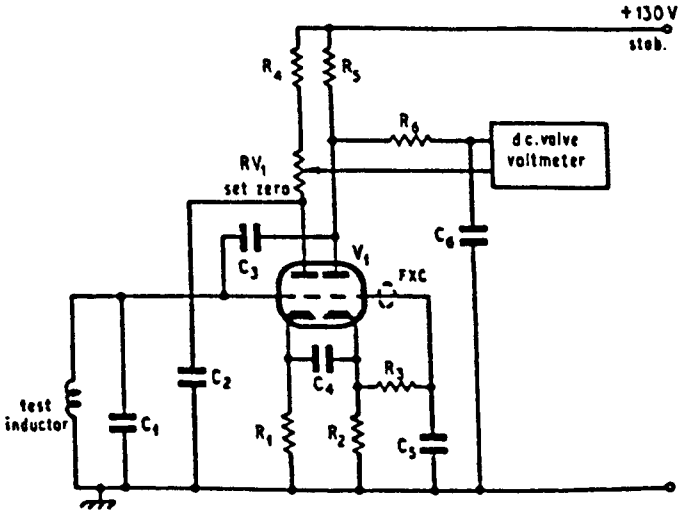


Figure 4.3(c) The marginal oscillator circuit due to Brice & Moore (1961) [9].

although Knight [7] describes a modification to his original oscillator which permits operation as a Robinson limited oscillator. This development was probably triggered by the first presentation, by Robinson a year earlier, of a valve version of the Robinson limited oscillator [10].

A few authors have reported the use of MOSFET's in their marginal oscillator designs. Yagi et al [11] describes a circuit for low temperature NMR work. This MOSFET operates well in this application as the circuitry is required to work at a temperature of 4K. Sullivan [12] describes a MOSFET circuit which closely resembles the original Pound & Knight valve design but points out that at moderate rf voltages, noise becomes a problem and a similar JFET design is then presented as a lower noise alternative. These MOSFET circuits are all operated as van der Pol oscillators.

McIver [13] describes a Robinson limited oscillator in which the amplifier nonlinear stage comprises a MOSFET source follower and justifies its use by pointing out the high input impedance obtained (which is true) and the low noise level (which is doubtful when compared with a similar JFET stage).

Most recent JFET marginal oscillator circuits have resembled the Pound & Knight configuration and four examples are reproduced as Figure 4.3(d)-(g). The circuits due to Viswanathan et al [14] and McElroy [15] make up one basic configuration with the source resistors separated by a capacitor whilst those due to Probst et al [16] and Idoine et al [17] are similar but have a single source resistor. The similarity of these circuits to the valve circuits of Figures 4.3(a)-(c) is worthy of note. A completely different JFET configuration is presented by Rahf [18] where only a single JFET is included.

JFETS exhibit relatively both low noise and high gate input impedances and appear to be the preferred solid state active element in Pound & Knight configuration van der Pol type marginal oscillators.

Integrated amplifiers produce limiting action by either saturation of some characteristic (often supply rail limiting) or through some 'accidental' nonlinearity. Robinson [3] points out that the lack of control over the the nonlinearity makes design of integrated amplifier based van der Pol type marginal oscillators difficult although limited oscillators may be implemented without this problem. Indeed, Conradi et al [19] who use a van der Pol integrated amplifier oscillator in an ultrasonic application find that an additional gain control loop is required to stabilise the circuit. A good integrated amplifier should be highly linear and although a van der Pol oscillator based on such an amplifier exhibits high signal sensitivity, it is also very noisy and the

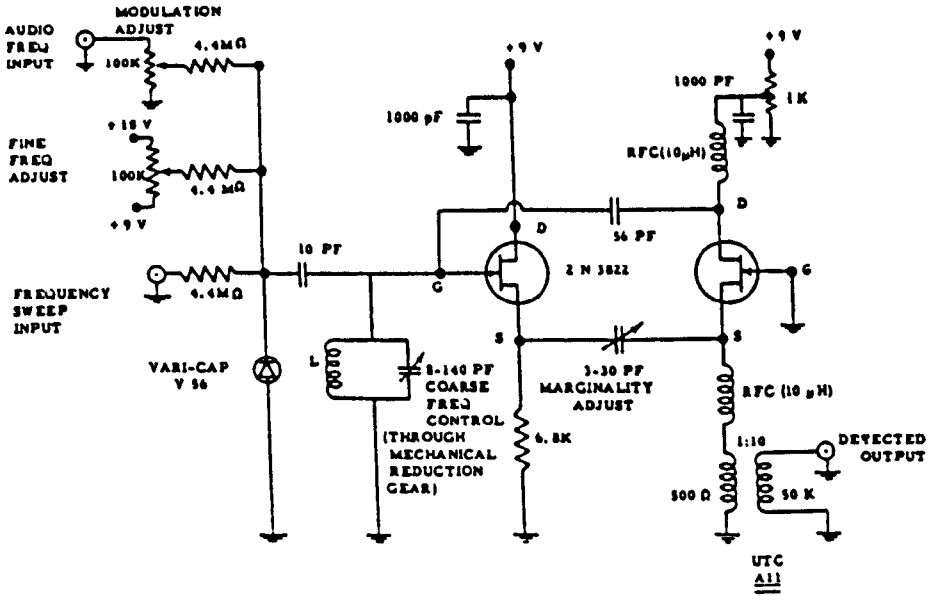


Figure 4.3(d) The marginal oscillator circuit due to Viswathan et al (1968) [14].

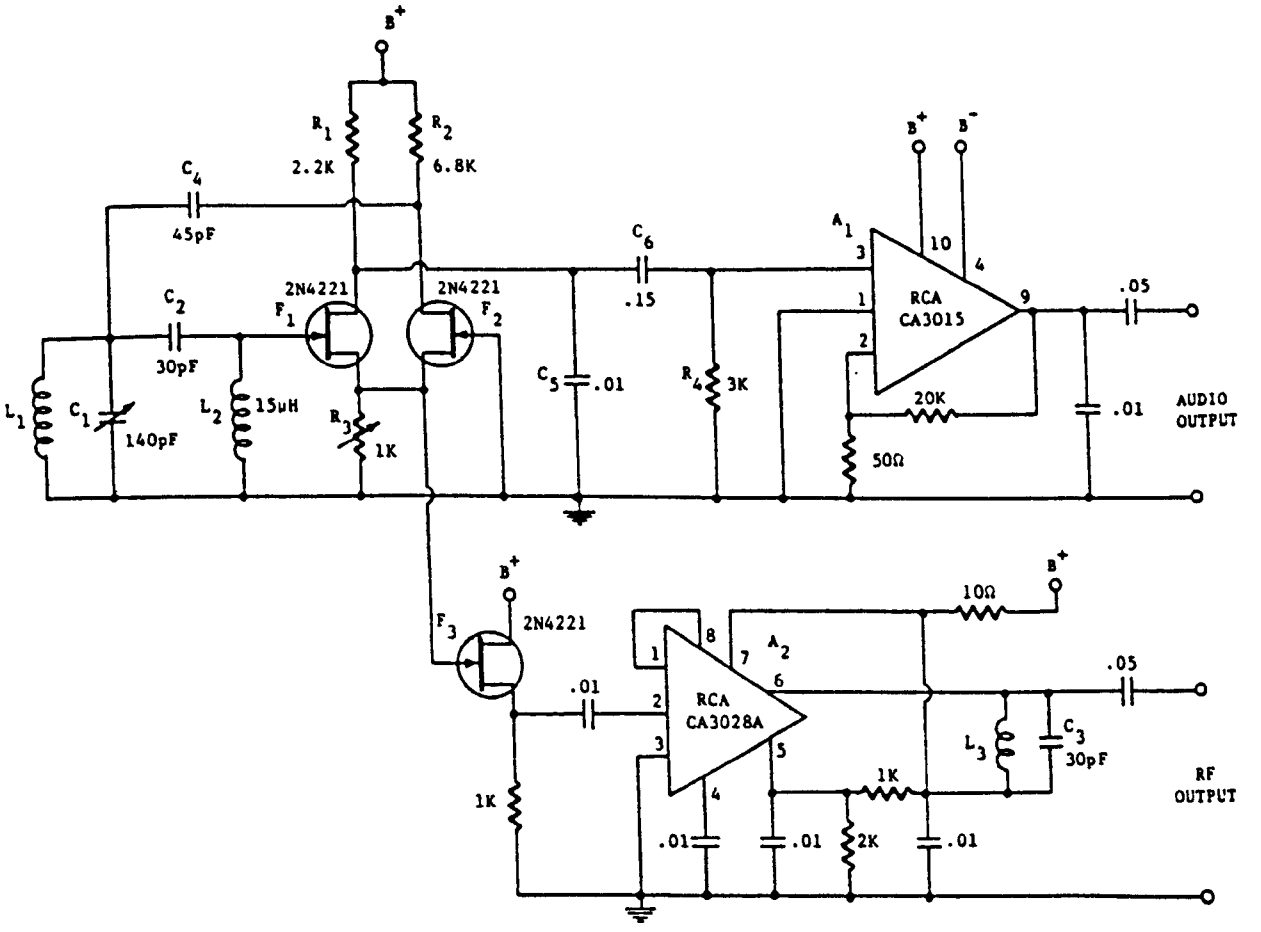


Figure 4.3(e) The marginal oscillator circuit due to Idoine et al (1971) [17].

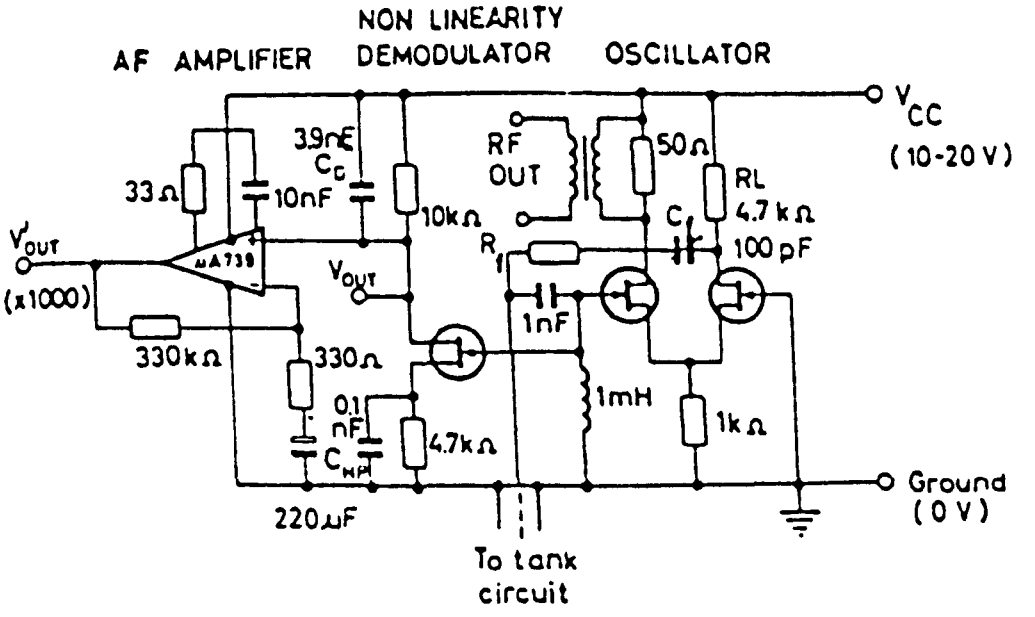


Figure 4.3(f) The marginal oscillator circuit due to Probst et al (1976) [16].

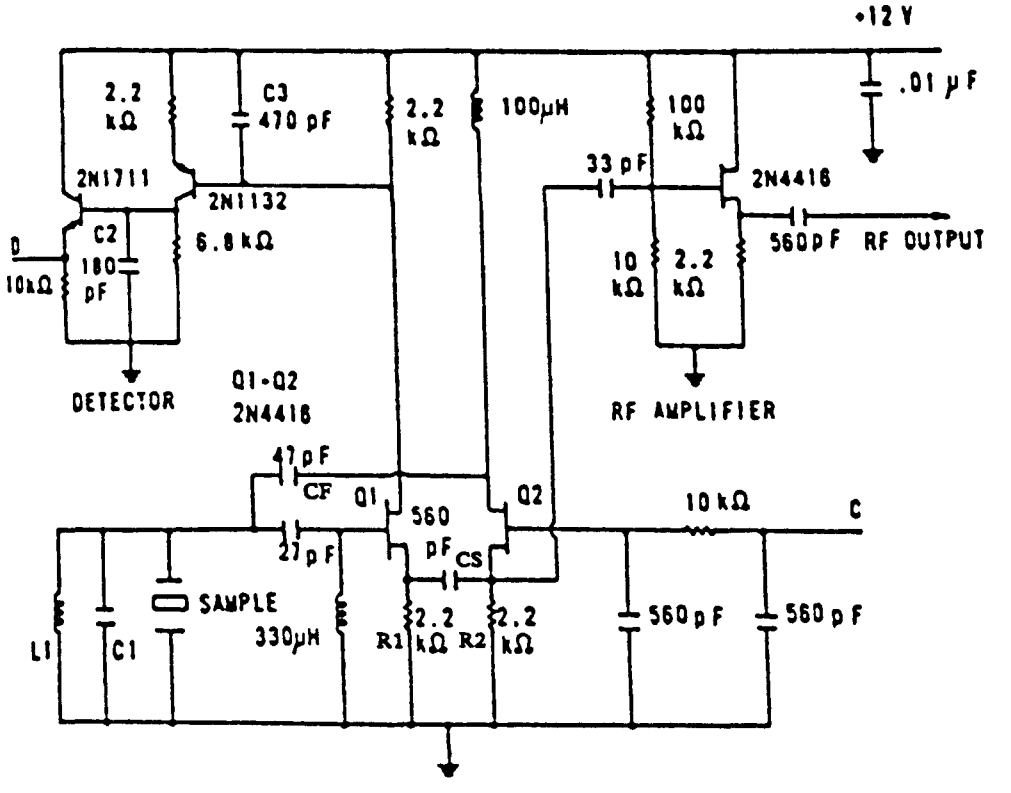


Figure 4.3(g) The marginal oscillator circuit due to McElroy et al (1980) [15].

quiescent operating signal level is sensitive to drift.

Marginal oscillator designs which do not fall into the above categories include the frequency sweeping marginal oscillator which is described by Kemper et al [20] and by Amano et al [21]. Such arrangements are, however, difficult to implement due to the variation of circuit parameter with changing frequency.

Another unusual approach is due to Deschamps et al [22] where a Robinson limited oscillator circuit is presented with both the amplifier and limiter stages being based on integrated circuit bipolar junction transistor arrays. The particular application makes this approach appropriate although the authors admit that the noise performance is not particularly good.

4.4 OSCILLATOR CIRCUITRY USED IN THIS WORK

Various oscillator configurations which have been used in spectrometer applications were reviewed in §4.3 and we now develop a design based on these whilst also taking into consideration the factors described in §4.2.

4.4.1 General Considerations

Desirable features of an oscillator for our application are:

- High signal to noise ratio for both frequency and amplitude perturbation effects.
- Good frequency and amplitude stability.
- Output signals easily obtainable.
- Minimal intermixing between frequency and amplitude perturbation effects.

It is generally recognised [14,17] that JFET's offer the best low-noise performance as the active components in the oscillator nonlinear amplifier due to their low flicker, shot and thermal noise characteristics.

Frequency stability is largely determined by tuned circuit characteristics and is not strongly dependent upon the nonlinear amplifier's characteristics (see §4.2). We cannot, therefore, influence frequency stability except by ensuring that the tuned circuit Q is maximised.

An oscillator circuit whose signal amplitude is sensitive to small tuned circuit conductance perturbations (i.e. $[G/V_o.dV_o/dG] \gg 1$) will of course respond significantly to changes in tuned circuit conductance due to other factors such as temperature change and component aging effects. The corresponding DC change in signal amplitude is not

usually important in a system where we wish to observe transient effects provided that these occur much faster than the background drift effects. However, when drift effects significantly alter the amplitude sensitivity of the system (i.e. the factor $[G/V_o.dV_d/dG]$ changes significantly) then problems can arise.

A high value of $[G/V_o.dV_d/dG]$ corresponds to operating with the nonlinear amplifier in a near-linear state. From §4.2 we note that the factor $[G/V_o.dV_d/dG]$ is not included in the expression for the system amplitude signal-to-noise ratio and it makes good sense, therefore, to reduce the factor as much as possible. The simplest method of achieving this is to operate at a sufficiently high level to ensure strong saturation effects within the nonlinear amplifier.

An output voltage signal must, of course, be available to permit subsequent signal processing and the only consistent principle used by previous workers is that of avoiding connection to the tuned circuit and, instead, connecting to some convenient node in the nonlinear amplifier. Provided that the buffering circuitry making the connection does not significantly load the oscillator (i.e. the buffer must have a relatively high input impedance at the operating frequency) then the selection of the node is not critical.

Marginal oscillators operate without interaction between amplitude and frequency perturbation effects only when no phase change (at the fundamental frequency of oscillation) occurs through the nonlinear amplifier [15]. It is therefore necessary that the phase shift be maintained as close to zero as possible.

4.4.2 Specific Design Considerations

In §4.3 we noted that the most popular marginal oscillator design of recent years has been the van der Pol twin-JFET type and, bearing in mind the factors introduced so far in this chapter, we felt that such a configuration would be suitable for this work. The basic circuit of McElroy [15] (see Figure 4.3(g)) was therefore adopted.

None of the authors who developed the circuits reproduced in §4.3 have attempted to analyse their designs. Such an analysis is difficult, particularly where JFET source connections are made using a capacitor, and this would possibly explain the omission. Indeed, the only attempt at analysing a real marginal oscillator circuit is due to Sullivan [12] where unfortunately he appears to obtain dimensionally incorrect equations.

An analytic description of the operation of the circuit due to McElroy is probably not obtainable but a qualitative explanation of its operation is feasible. The voltage

signal across the tuned circuit is fed, via the high-pass LC filter, to the gate of Q1. The high pass filter primarily suppresses 50Hz mains interference.

The combination of Q1 and source resistors R1 and R2 forms a source-follower amplifier whose key features are high input impedance, low output impedance and a voltage gain of just less than unity for small signals. The voltage gain of the stage becomes progressively lower with increasing signal level and hence contributes to the overall nonlinear nature of the amplifier.

The output of the source-follower stage is fed into a common-gate amplifier via capacitor CS which effectively presents a short circuit at signal frequencies. The common-gate stage provides substantial voltage gain and the output signal at the drain of Q2 is fed back to the tuned circuit via capacitor CF to a buffer amplifier for subsequent signal processing. As the signal level is raised, the common-gate stage also becomes increasingly nonlinear and contributes to the overall amplifier nonlinearity.

Potentiometers may be used to implement resistors R1 and R2 as these provide flexibility in setting oscillator signal levels whilst all other components remain fixed. The potentiometers may be subsequently locked at specific values.

The marginal oscillator circuit used is illustrated in Figure 4.4.2(a); the value of the parallel tuned circuit capacitance is not specified as it comprises, to a high degree, the capacitance of the inductive sensor connecting cable. Also, at this stage, the tuned circuit inductance is not specified as it depends on the coil geometry selected (see §6.1.1).

4.5 SUMMARY AND CONCLUSIONS FOR CHAPTER FOUR

This chapter is summarised and concluded as follows:

- The most appropriate method for detecting perturbations in the impedance of an inductive transducer is by incorporating the transducer into a marginal oscillator circuit.
- In the marginal oscillator, the coil transducer forms the inductance of the parallel resonant tuned circuit driven by a feedback amplifier.
- Perturbations in coil series resistance (or, equivalently, the parallel tuned circuit conductance) amplitude modulate the oscillator signal, whilst perturbations in coil inductance cause frequency modulation. To a good approximation, the effects do not interact provided that the feedback amplifier does not produce significant phase shift at the operating frequency.

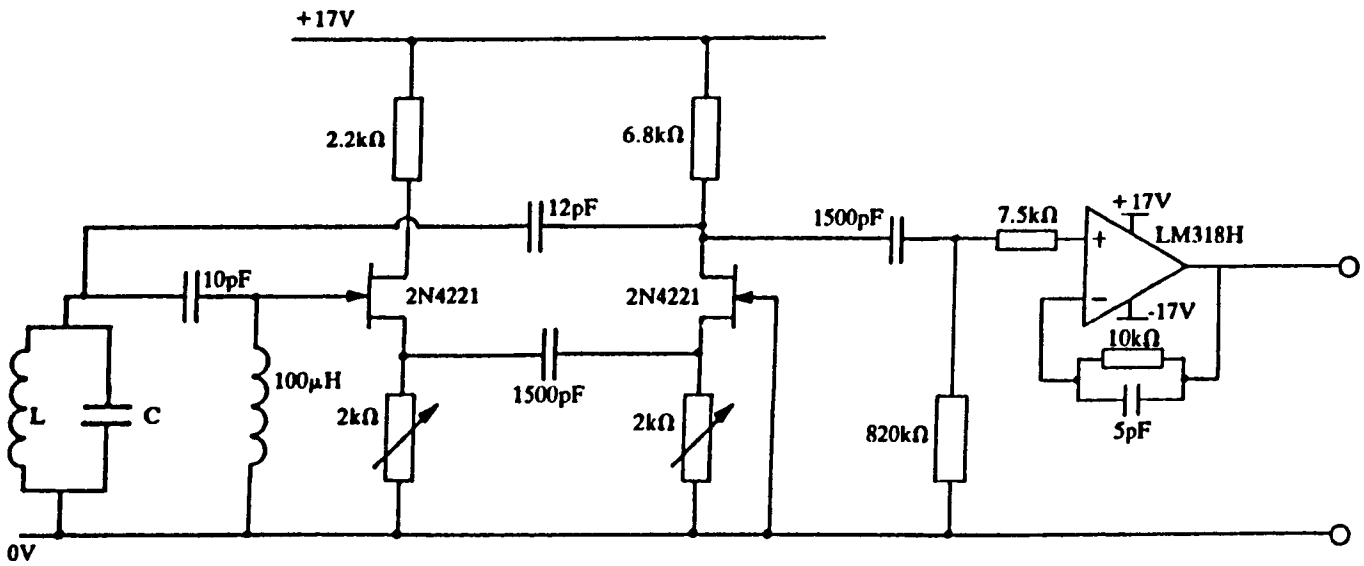


Figure 4.4.2(a) The marginal oscillator circuit used in this work. The circuit was designed for operation at ~10MHz and a suitable unity-gain buffer was added as shown to enable connection to other circuitry.

- The signal equations are given by the approximations

$$\frac{\Delta V_o}{V_o} = \left[\frac{G}{V_o}.\frac{dV_o}{dG}\right]\frac{\Delta G}{G}$$

and

$$\frac{\Delta f}{f} = \left(-\frac{1}{2}\right)\frac{\Delta L}{L}$$

- The noise equations are given by

$$V_{n\,rms} = \frac{1}{G}\left(\frac{G}{V_o}.\frac{dV_o}{dG}\right)\frac{I_{n\,rms}}{2^{1/2}}$$

and

$$f_{n\,rms} = \frac{1}{4\pi}\left[\frac{1}{2^{\frac{1}{2}}\,C\,V_o}\right]I_{n\,rms}$$

- The oscillator response to signal and noise perturbations is a low-pass filtered effect.
- The amplitude signal and noise equations include the term $\left[\frac{G}{V_o}.\frac{dV_o}{dG}\right]$ which is a function of the feedback amplifier non-linearity at the quiescent operating voltage. When this term is large, both signal and noise effects are enhanced by the same factor.
- A marginal oscillator with the factor $\left[\frac{G}{V_o}.\frac{dV_o}{dG}\right] = -1$ is termed a Robinson limited oscillator whilst a more sensitive oscillator, with the factor $\left[\frac{G}{V_o}.\frac{dV_o}{dG}\right] < -1$, is termed a Van der Pol oscillator.
- Many marginal oscillator circuits using an assortment of active component types have been presented in the literature and the commonly-used JFET source-follower / common-gate configuration has been selected for this work. Such a scheme allows high signal-to-noise ratios as well as easy access to the signal for subsequent signal processing.

CHAPTER FOUR: REFERENCES

1. Gregory B A: 'An introduction to electrical instrumentation and measurement systems', 2nd ed, Macmillan, London (1981), Chapter 3.

2. Shepherd J, Morton A H & Spence L F: 'Higher Electrical Engineering', 2nd ed, Pitman, London (1970), Chapter 1.

3. Robinson F N H: 'Noise in oscillators', Int J Electronics, **56** (1984) 63-70.

4. Hughes D G and Smith M R: 'On the nuclear magnetic resonance detection characteristics of marginal and Robinson oscillators', J Phys E, 4 (1971) 13-20.

5. Leeson D B: 'A simple model of feedback oscillator noise spectrum', Proc IEEE, **54** (1966) 329-30.

6. Pound R V and Knight W D: 'A radiofrequency spectrograph and simple magnetic-field meter', Rev Sci Instrum **21** (1950) 219-25.

7. Buss L and Bogart L: 'Wide range marginal oscillator for operating nuclear resonance probes through flexible cable', Rev Sci Instrum, **31** (1960) 204-5

8. Knight W D: 'High level oscillator for nuclear resonance', Rev Sci Instrum, 32 (1961) 95.

9. Brice J C and Moore P: 'Contactless resistivity meter for semiconductors', J Sci Instrum, **38** (1961) 307.

10. Robinson F N H: 'Nuclear resonance absorption circuit', J Sci Instrum, **36** (1959) 481-487.

11. Yagi H, Inoue M, Naito T and Tatsukawa T: 'NMR marginal oscillator with MOS-FET operating at low temperature', Jap J Appl Phys, **12** (1973) 1794-98.

12. Sullivan N: 'Nuclear resonance spectrometers using field effect transistors', Rev Sci Instrum, **42** (1971) 462-65.

13. McIver R T: 'A solid-state marginal oscillator for pulsed ion cyclotron resonance spectroscopy', Rev Sci Instrum, **44** (1973) 1071-74.

14. Viswanathan T L, Viswanathan T R, Sane K V: 'NQR spectrometer using field effect transistors', Rev Sci Instrum, **39** (1968) 472-75.

15. McElroy R G C: 'Use of a capacitively coupled marginal oscillator for contactless transient conductivity measurements', Rev Sci Instrum, **51** (1980) 1374-77.

16. Probst P A, Collet B & Macinnes W M: 'Marginal oscillator optimised for radiofrequency size effect measurements', Rev Sci Instrum, **47** (1976) 1522-26.

17. Idoine J D and Brandenberger J R: 'FET marginal oscillator circuit', Rev Sci Instrum, **42** (1971) 715-717.

18. Rahf L: 'New sensitive marginal oscillator', Rev Sci Instrum, **52** (1981) 1361-63.

19. Conradi M S & Miller J G: 'A transmission oscillator ultrasonic spectrometer', Rev Sci Instrum, **45** (1974) 358-60

20. Kemper P R & Bowers M T: 'Frequency-scanning marginal oscillator for ion cyclotron resonance spectroscopy', Rev Sci Instrum, **53** 989-96.

21. Amano C, Goto Y & Inoue M: 'Simple wideband frequency-swept marginal oscillator detector for ion-cyclotron resonance spectrometers', Int J Mass Spectrum Ion Phys, **32** (1979) 67-76.

22. Deschamps P, Vaissiere J & Sullivan N S: 'Integrated circuit Robinson oscillator for NMR detection', Rev Sci Instrum, **48** (1977) 664-68.

CHAPTER FIVE

DEMODULATION TECHNIQUES

In the previous two chapters we have described the background and theory of the transducer and marginal oscillator stages of our system and in this chapter we deal with the subsequent demodulation stages. As stated in §4.2.1, amplitude and frequency demodulation of the marginal oscillator signal yields a measure of the perturbation of the inductive transducer resistance and inductance due to a metal particle passing through. A range of demodulation techniques is available and we shall, in the following sections, review frequency and amplitude demodulation techniques generally and the techniques chosen for this work in particular.

5.1 FM DEMODULATION

FM (or frequency) demodulation is the process by which frequency changes relative to a carrier frequency are converted into voltage signals. The circuitry used to achieve this is known as an FM demodulator. Desirable features of an FM demodulator include:

- High linearity
- Low Noise
- Sufficiently high frequency response for the application

To obtain FM demodulation we require some form of circuitry producing an output amplitude signal linearly dependent on the input frequency. Such a scheme may be characterised by a transfer function as shown in Figure 5.1(a). Provided that the carrier frequency, f_c, is much greater than the peak modulating frequency, f_m, it is usual effectively to reduce the carrier frequency to an intermediate frequency, f_i. This *frequency down-conversion* process makes subsequent frequency-to-voltage conversion more effective. To illustrate this point consider two alternative FM demodulator schemes as shown in Figure 5.1(b). In the first scheme we attempt to demodulate *at the carrier frequency* and, as a consequence, the ratio f_m/f_c is comparatively small. Moreover, it is often difficult to design a frequency to voltage converter to operate at the carrier frequency.

In the second scheme we initially perform a down-conversion and subsequently demodulate *at the intermediate frequency*. At this stage we are detecting a peak frequency change, f_m, about the relatively low intermediate frequency 'carrier'. To ensure proper operation of the scheme, the intermediate frequency must be

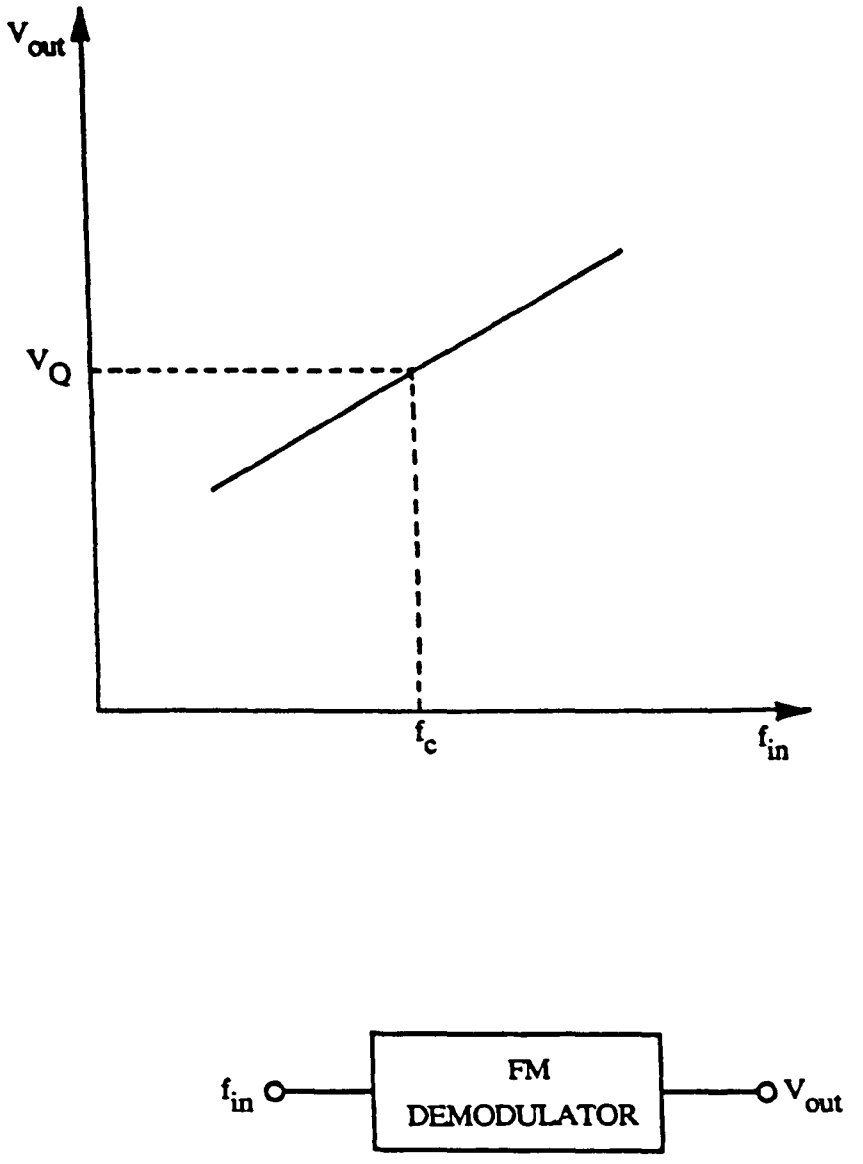
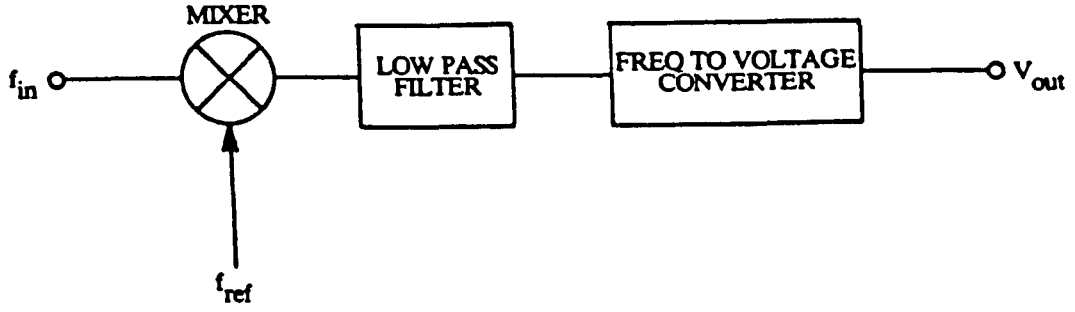


Figure 5.1(a) The transfer characteristic of an FM demodulator about the quiescent operating point, (f_c, V_Q).



(i)

(ii)

Figure 5.1(b) Block diagrams of two alternative FM demodulation schemes; (i) with downconversion prior to the frequency to voltage conversion stage, and (ii) direct frequency to voltage conversion at the carrier frequency.

considerably higher than the highest modulating frequency.

Many frequency-to-voltage conversion circuits respond to the *relative change* in frequency due to modulation and consequently the second scheme provides an intrinsic signal gain compared to the first. It also provides an intrinsic noise gain but because we require a lower gain amplifier than for the first scheme, amplifier noise effects are less significant and generally ensures a better output signal-to-noise ratio.

Several specific FM demodulator techniques have been developed. The earlier of these depend on circuitry whose amplitude transfer characteristic is highly frequency dependent in the frequency range about the carrier (or intermediate frequency if a down-conversion process has been effected). Amplitude detection is then required to complete the demodulation process. Examples of such demodulators are the Foster-Seeley Discriminator [1] and the Ratio Detector, as reproduced in Figure 5.1(c).

An alternative approach is the use of a phase locked loop [1] where an electronic servo-loop provides an output voltage linearly dependent on the input frequency and this is shown in block diagram form in Figure 5.1(d).

A third approach, which has become feasible since the introduction of digital integrated circuits, is pulse demodulation [2]. Two basic options exist and these are shown schematically in Figure 5.1(e). The first scheme depends on the positive-going edge of the input signal triggering a counter which is clocked by a separate signal source whose frequency is much greater than that of the carrier. On the negative edge of the input signal, the counter is stopped and the accumulated binary value fed into the digital-to-analogue converter which, in turn, outputs a voltage signal inversely proportional to the input frequency.

The second pulse demodulator scheme depends on the monostable generating a constant pulse length in reaction to a positive edge on the input signal. The monostable output is then low-pass filtered to produce a signal inversely proportional to the input signal.

With the availability of modern digital circuitry, implementing a frequency down-converter followed by a monostable-based pulse demodulator is reasonably straightforward and such a combination has been selected for FM demodulation in this work.

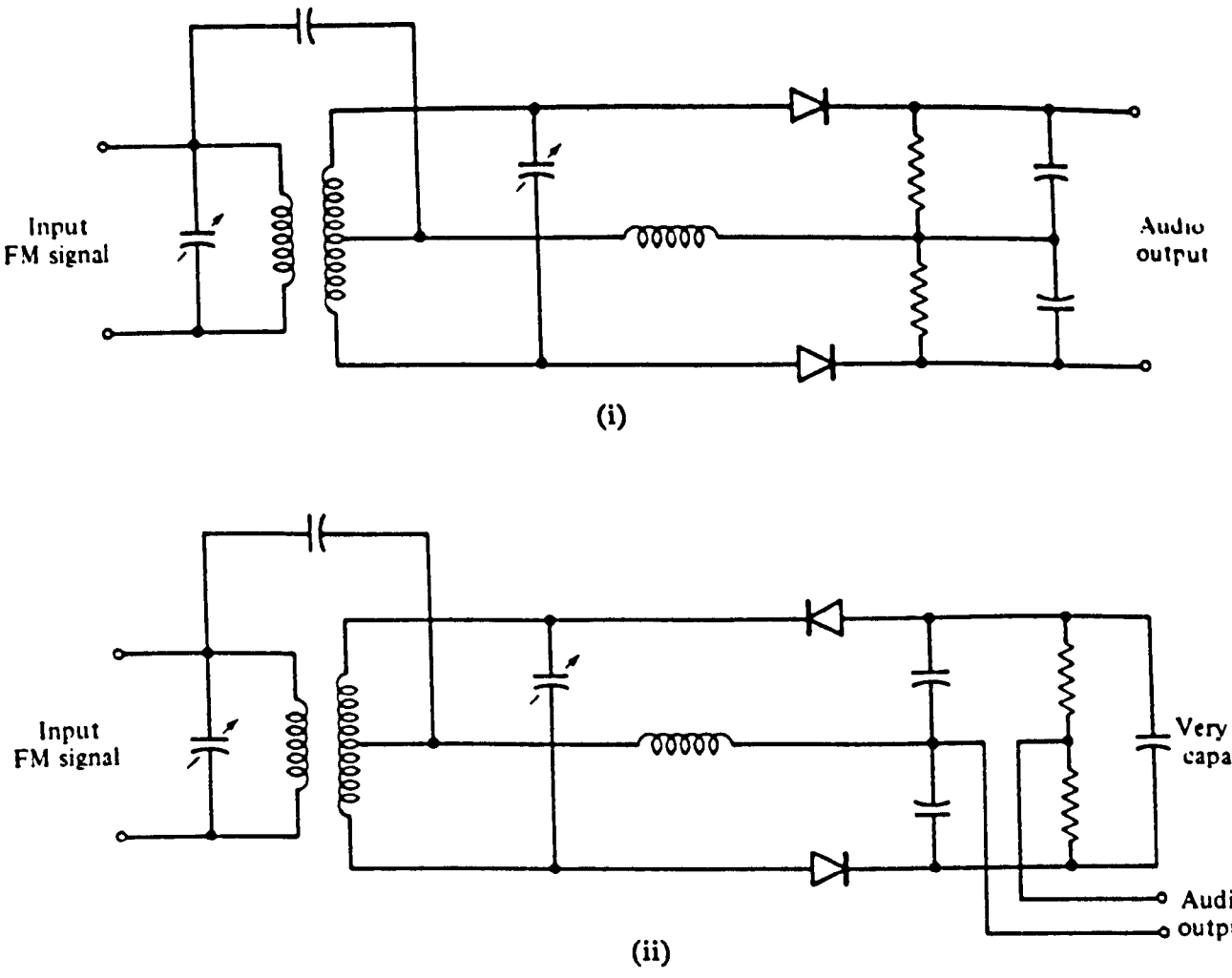


Figure 5.1(c) FM demodulator types: (i) The Foster-Seeley discriminator; (ii) The ratio detector [1].

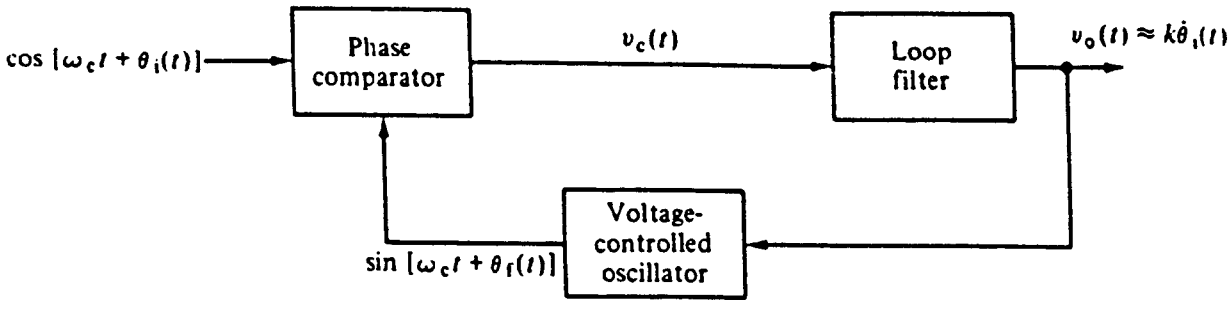


Figure 5.1(d) The phase locked loop [1].

(i)

(ii)

Figure 5.1(e) Block diagrams of pulse demodulator schemes; (i) the counter based scheme, and (ii) the monostable based scheme [2].

5.2 PRESENT FM DEMODULATION CIRCUIT

A suitable choice of FM demodulator for this work is the frequency conversion / monostable-based pulse demodulator scheme and we describe the circuitry implemented in this section.

5.2.1 Block Diagram Description

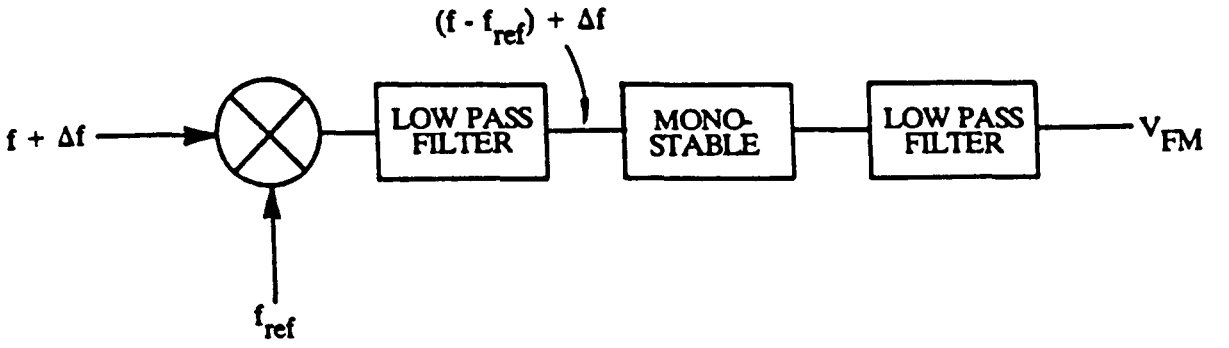
The frequency down-conversion process (see Figure 5.2.1(a)) is implemented using a mixer circuit which produces sum and difference frequency terms which are fed through a low-pass filter to produce the difference term. The signal is then fed into a monostable multivibrator circuit to generate a waveform whose mark-to-space ratio is proportional to the intermediate frequency. This signal is, in turn, low-pass filtered to produce an output signal inversely proportional to the intermediate frequency. This is the demodulated signal.

The voltage to frequency conversion gain of the demodulator system, G_{FM} (units V/Hz), is equal to the gradient of the f_{in}-V_{out} characteristic in Figure 5.2.1(a). In addition, the final low-pass filter, used to extract the dc component from the monostable output signal limits the upper frequency response of the system to frequency perturbations.

5.2.2 Circuit Diagram Description

The circuit diagrams for the FM demodulator adopted are shown in Figures 5.2.2(a) and (b). The input signal is fed, via an emitter follower buffer, into a schmitt trigger inverter. The output of the inverter is fed, in parallel with the signal from a crystal oscillator module, into an exclusive-or gate, the output of which contains a large component at the *intermediate frequency*. After low-pass filtering, the signal is fed into the LM311-based comparator circuit whose output is a TTL compatible square waveform. This completes the frequency downconversion process.

The intermediate frequency signal is then fed into the monostable which is adjusted, using the 20kΩ potentiometer, to produce an output with an approximately unity mark-to-space ratio. This is, in turn fed to the LM318-based stage which generates a bipolar intermediate frequency square waveform which can be attenuated using the 5kΩ potentiometer. This signal is subsequently fed into the active low-pass filter circuit (Figure 5.2.2(b)), whose output represents the FM demodulated signal, and has a low-pass 3dB cut-off frequency of ~5kHz.



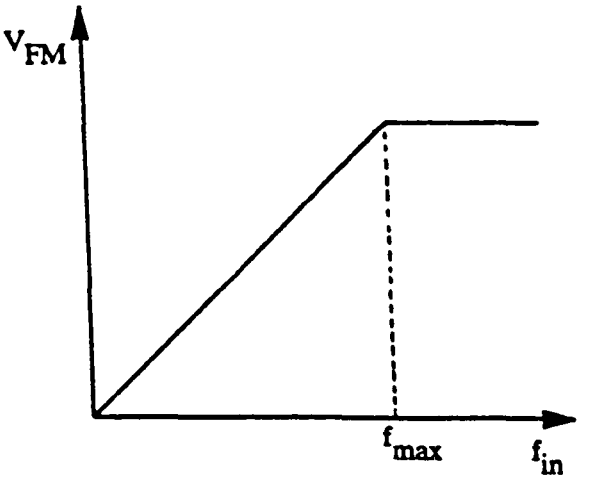


Figure 5.2.1(a) Block diagram of the frequency conversion / monostable-based pulse demodulator and the associated f-V transfer characteristic.

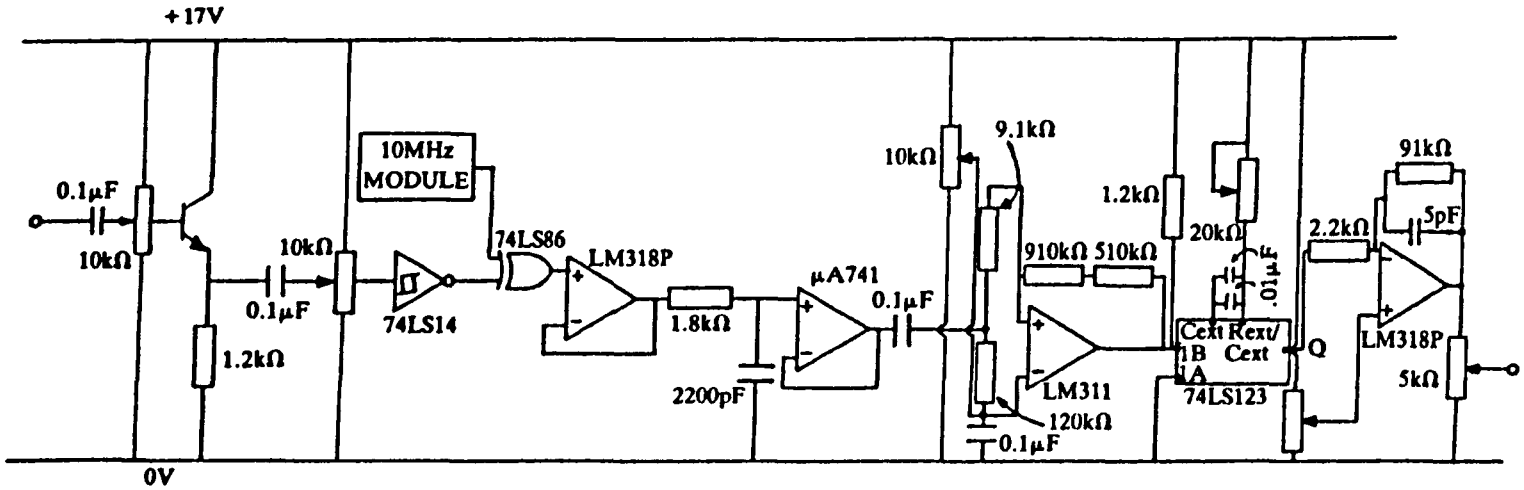


Figure 5.2.2(a) The initial stages of the FM demodulation system used in this work.

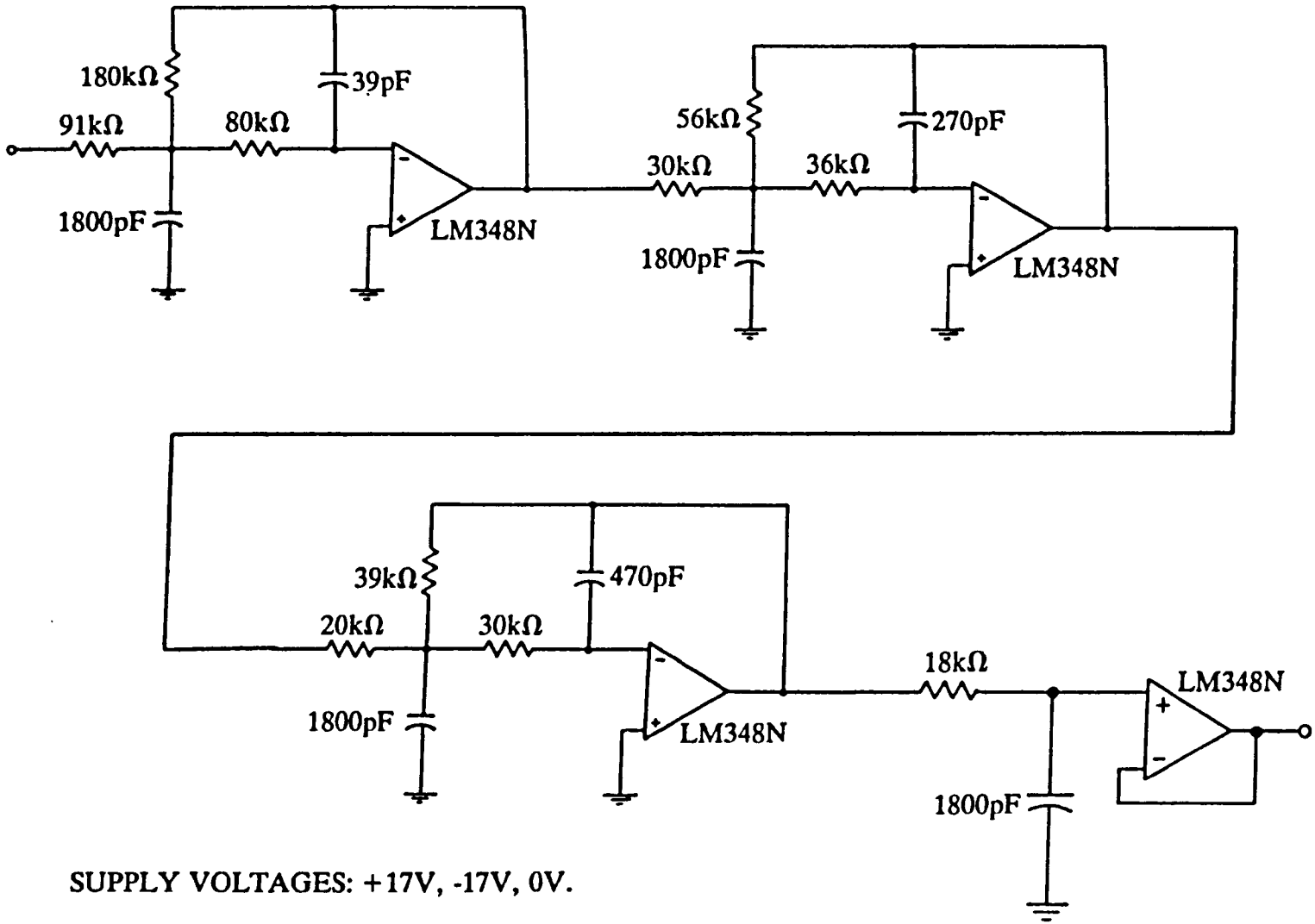


Figure 5.2.2(b) The final low-pass filter stage of the FM demodulation system used in this work. The circuit comprises a series of three Butterworth 2nd order stages cascaded with a simple first order stage. (From: Johnson D E *et al:* 'A handbook of active filters', Prentice-Hall, NJ USA, 1980.)

5.3 AM DEMODULATION

AM demodulation (or detection) may be defined as the process by which the *amplitude* of a high frequency AC signal is converted into a new, relatively low frequency signal. AM demodulators are based on a variety of principles and in this section we review the most common of these. We may identify two major groupings: the coherent and the non-coherent demodulators.

5.3.1 Coherent AM Demodulation

Coherent AM demodulation techniques depend upon accurately obtaining the *phase* of the carrier and this process is usually achieved using a phase locked loop [3] as shown in Figure 5.3.1(a). Such a technique is known to provide a signal-to-noise ratio improvement over non-coherent methods, particularly for low signal-to-noise ratios [4] but the relative complexity of the associated circuitry has prevented widespread application.

5.3.2 Non-coherent AM Demodulation

Non-coherent AM demodulation techniques do not depend upon the phase of the incoming carrier signal and comprise three main techniques: square-law, rectifier and envelope detectors.

Square-law detection [5] (Figure 5.3.2(a)) depends on the carrier and modulating signals generating products including a component at the baseband frequency which is obtained by low-pass filtering the output. Square-law detection is infrequently used due to the low level of the output signal.

The rectifier detector [1] (Figure 5.3.2(b)) produces an output whose average value, obtained by subsequent low-pass filtering, represents the demodulated signal. The carrier frequency must be substantially higher than the highest modulating frequency in order that the carrier can be adequately filtered out. Use of this type of demodulator is normally restricted to circumstances where the slow exponential discharge rate of the envelope detector (see below) is not acceptable, such as in high speed pulse detection.

The envelope detector [1] is the simplest and most commonly used AM demodulator scheme and is shown in Figure 5.3.2(c). The envelope is obtained by charging the RC combination via the diode and the influence of the RC time constant on the process is also illustrated in Figure 5.3.2(c). The resulting demodulated signal is usually passed through a low-pass filter to remove any residual carrier frequency

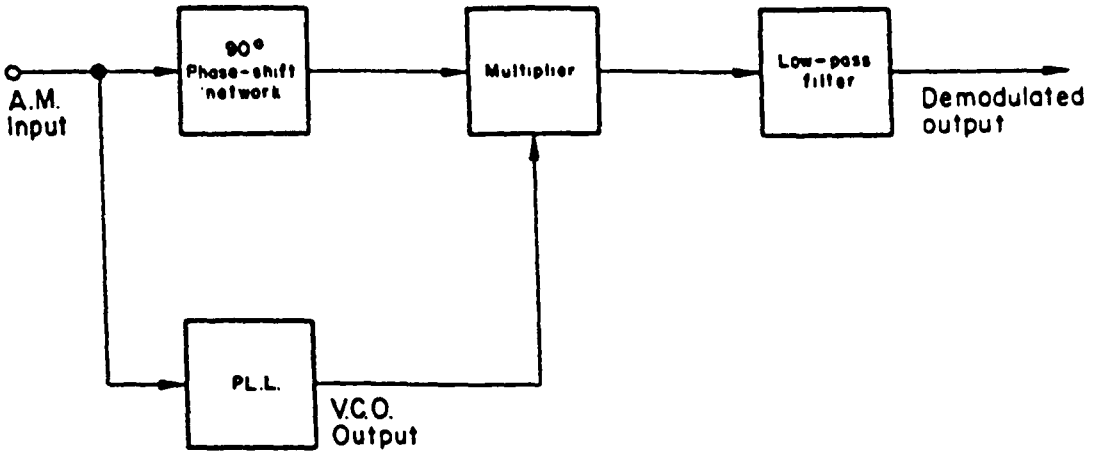


Figure 5.3.1(a) Coherent AM demodulation using a phase locked loop [3].

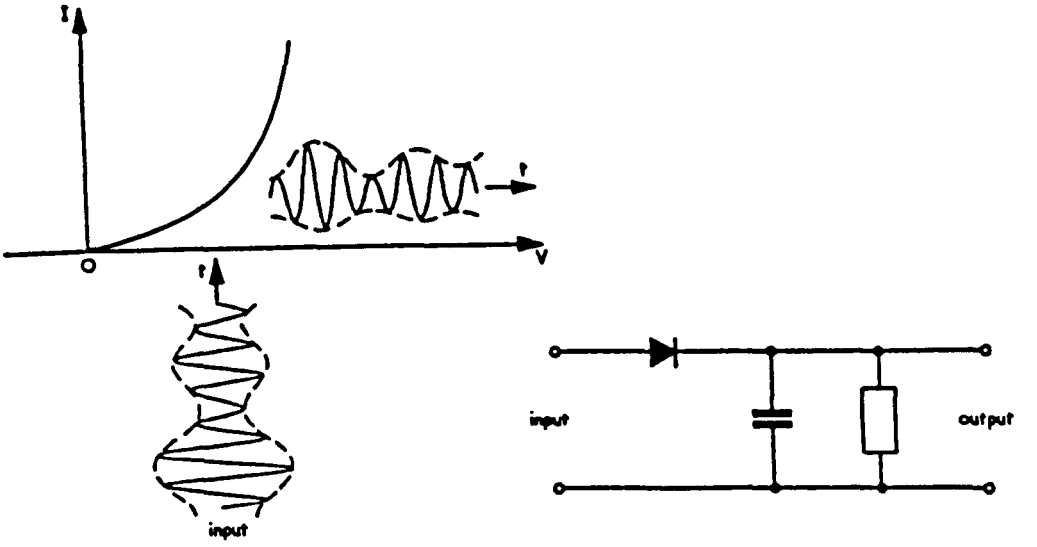


Figure 5.3.2(a) The square-law detector [5].

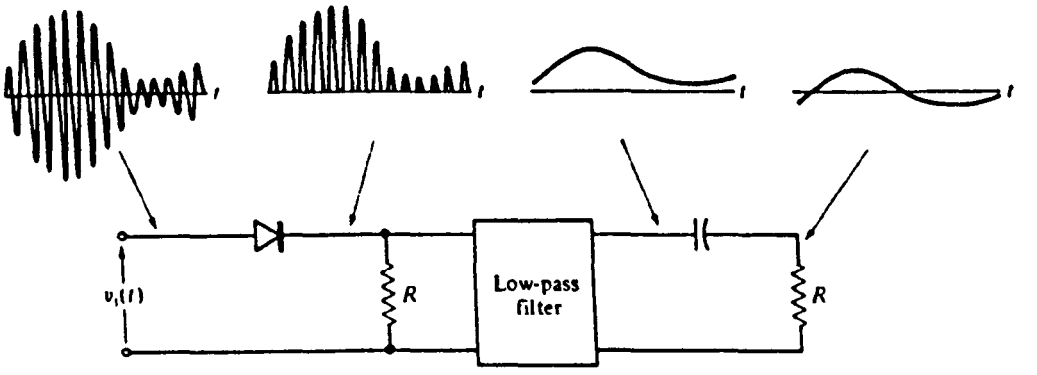


Figure 5.3.2(b) The rectifier detector [1].

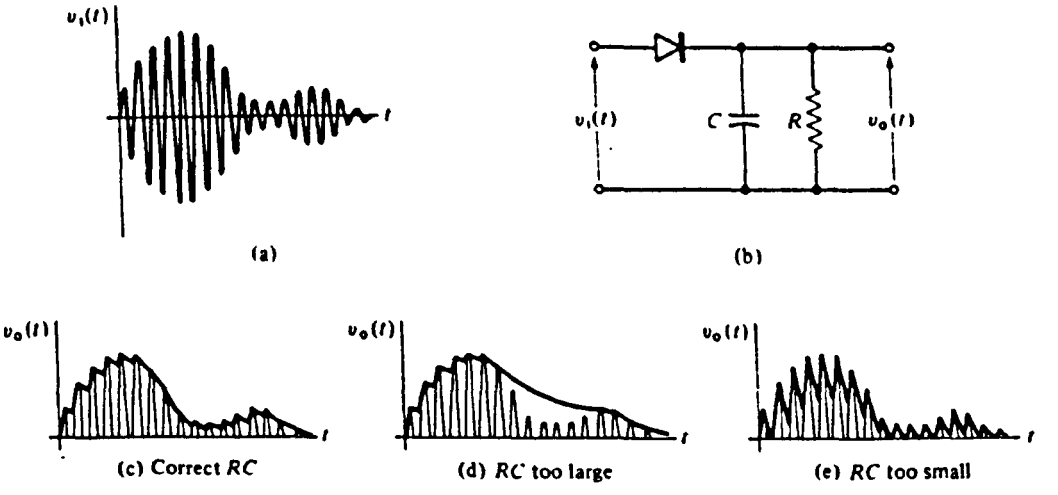


Figure 5.3.2(c) The envelope detector [1].

component. Provided that the carrier frequency is much greater than the modulating frequencies, envelope detection is a cheap and effective option.

5.4 PRESENT AM DEMODULATION CIRCUIT

We realised that AM demodulation performed at the original carrier frequency would obviate the need to downconvert to an intermediate frequency, thereby simplifying the required circuitry. As our carrier frequency was approximately 10MHz, the use of a phase locked loop coherent demodulation technique would have been difficult and as a consequence one of the possible non-coherent techniques appeared more suitable. The square-law and rectifier detectors were considered but we selected the envelope detector in preference due to its inherent simplicity.

5.4.1 Block Diagram Description

A block diagram of the envelope demodulator implemented in this work is shown in Figure 5.4.1(a). Initially, an amplifier ensures that the signal level is sufficient to drive the envelope detector stage effectively. The envelope detector stage is, in turn, followed by a low-pass filter stage.

The AM demodulator system gain, G_{AM} (dimensionless), is determined by the product of the gains of the pre-amplifier and the post-amplifier. Also the frequency response of the system to amplitude perturbations is limited by the bandwidth of the final low-pass filter and/or the effective bandwidth, at the quiescent operating level, of the envelope detector.

5.4.2 Circuit Diagram Description

The circuit diagram of the amplitude demodulator is shown in Figure 5.4.2(a). The first stage comprises a source follower buffer stage, after which the signal is amplified using the bipolar transistor cascode amplifier and fed through an emitter follower stage. This is followed by the envelope detector which comprises a germanium signal diode with a resistor/capacitor parallel combination. The resulting signal is then AC coupled, amplified, low-pass filtered and AC coupled again to produce the amplitude demodulated signal.

The dominant low-pass filtering effect within the AM demodulator is due to the slew-rate limiting action of the envelope detector which has a time constant of ~0.6ms. For small amplitude perturbations, a low-pass filter cut-off frequency of ~1kHz can be attributed to the stage.

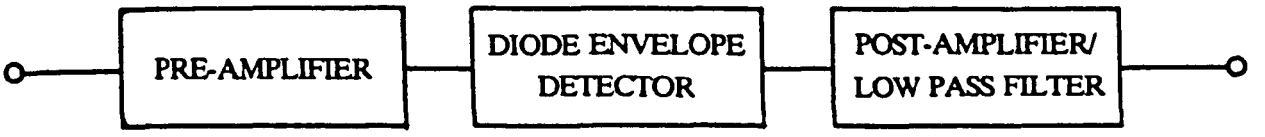


Figure 5.4.1(a) Block diagram of the envelope detector used in this work.

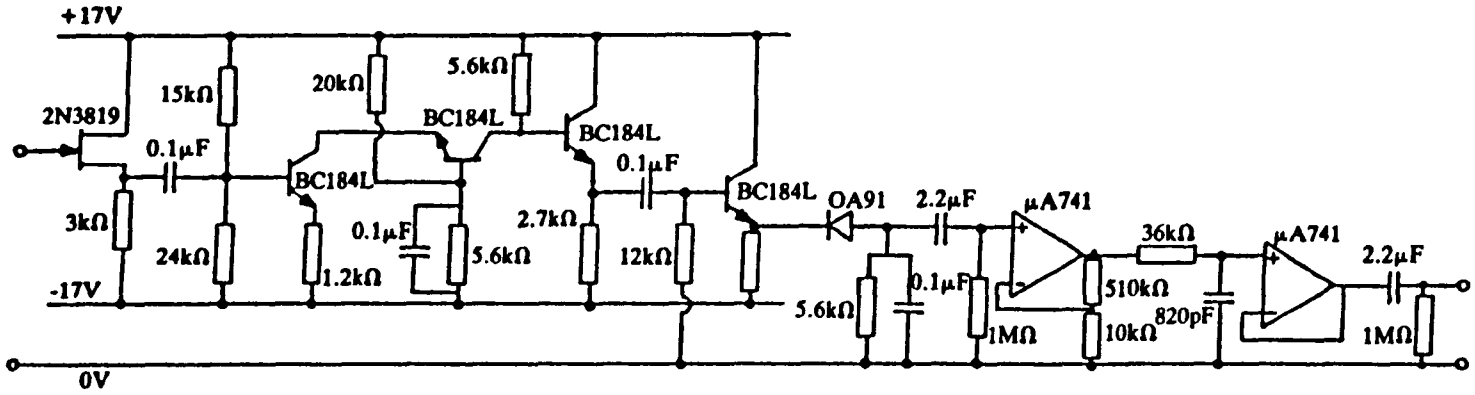


Figure 5.4.2(a) Amplitude demodulation system used in this work.

5.5 SUMMARY AND CONCLUSIONS FOR CHAPTER FIVE

This chapter is summarised and concluded as follows:

- To measure frequency and amplitude perturbations in the marginal oscillator we require both an FM and AM demodulator in the system
- FM demodulation can be effected using a variety of techniques. For this work we have implemented a scheme consisting of a frequency downconversion stage followed by a monostable-based pulse demodulator.
- AM demodulation can also be effected using a variety of different techniques. For this work we have implemented a scheme based on an envelope detector.

CHAPTER FIVE: REFERENCES

1. Stremler F G: 'Introduction to communication systems', Addison-Wesley, Reading Massachusetts, 2nd ed (1982).

2. Smith J: 'Modern Communication Circuits', McGraw-Hill, New York (1986).

3. Clayton G B: 'Linear Integrated Circuit Applications', Macmillan, London (1975).

4. Gardner F M: 'Phaselock Techniques', Wiley, New York, 2nd ed (1979), Chapter 9.

5. Betts J A: 'Signal Processing, Modulation and Noise', Hodder & Stoughton, (1970) Chapter 3.

CHAPTER SIX

THE OVERALL SYSTEM

In this chapter we combine the work of Chapters 3, 4 and 5 to obtain theoretical expressions for the FM and AM characteristics of the overall system. In addition, we obtain numerical estimates of the overall FM and AM responses of the system.

6.1 OVERALL SYSTEM RESPONSE TO A PARTICLE IN THE TRANSDUCER

The overall system response depends on the combined characteristics of the inductive transducer, the marginal oscillator and the FM and AM demodulation circuitry. We now need to review the key features of the individual system components to enable expressions for the overall response to be obtained. A block diagram of the complete system is shown in Figure 6.1(a).

6.1.1 The Inductive Transducer in the System

The results which relate the real and imaginary components of the D value of a given particle to the impedance change reflected across the coil transducer terminals are, from §3.4,

$$\frac{\Delta L}{L} = K_c \mathrm{Re}[D] \qquad (1)$$

and

$$\frac{\Delta R}{\omega L} = -K_c \mathrm{Im}[D] \qquad (2)$$

The value of K_c depends on the particle position within the coil as well as coil geometry and is a function of time for a particle flowing through the coil. The value $K_c(0,0)$ is of particular interest as it corresponds to the peak of the pulse perturbation generated by a particle moving along the coil axis and is denoted K_{co}.

The value of K_{co} is a function of coil geometry: the smaller the coil the larger the value. For this work, a coil of dimensions, length $\lambda = 6$mm and radius $A = 3$mm was selected.

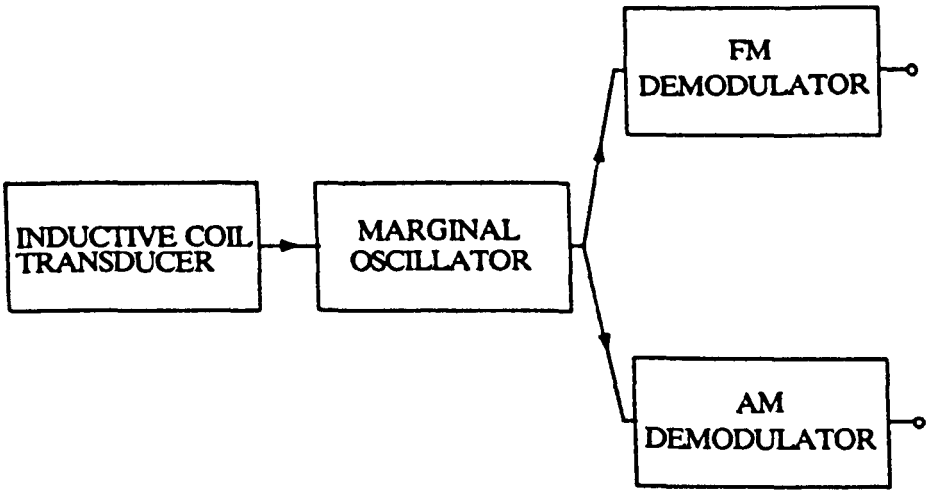


Figure 6.1(a) A block diagram of the complete system.

6.1.2 The Marginal Oscillator in the System

The marginal oscillator may be characterised by its frequency and amplitude perturbations due to changes in coil transducer resistance and inductance. For amplitude perturbations we may define a dimensionless constant M such that

$$\frac{\Delta V_o}{V_o} = M\frac{\Delta R}{\omega L} \tag{1}$$

Provided that the quality factor of the tuned circuit, $Q >> 1$

$$\frac{\Delta R}{R} = \frac{\Delta G}{G} \tag{2}$$

Combining Equations (1) and (2) yields

$$\frac{\Delta V_o}{V_o} = \left(\frac{M}{Q}\right)\frac{\Delta G}{G} \tag{3}$$

Comparing Equation (3) with the differential approximation

$$\frac{\Delta V_o}{V_o} = \left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right)\frac{\Delta G}{G} \tag{4}$$

yields

$$M = \left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right) Q \tag{5}$$

For frequency perturbations, from §4.2.3

$$\frac{\Delta f}{f} = \left(-\frac{1}{2}\right)\frac{\Delta L}{L} \tag{6}$$

The above equations are based on the assumption, which can be applied to our system, that we operate well below the low-pass cut-off frequencies associated with signal perturbation effects (see §4.2.1 and §4.2.3).

6.1.3 The Demodulators in the System

The key features of the FM and AM demodulators are

(i) The constant of proportionality, G_{FM}, between the frequency change and the corresponding FM demodulator output voltage change.

(ii) The constant of proportionality, G_{AM}, between the amplitude change and the corresponding AM demodulator output voltage change.

These effects may be written

$$V_{FM} = G_{FM}\,\Delta f \tag{1}$$

and

$$V_{AM} = G_{AM}\,\Delta V_o \tag{2}$$

where G_{FM} has dimensions V/Hz and G_{AM} is dimensionless.

The above equations apply only at frequencies below the demodulator effective low-pass cut-off frequencies (see §5.2.2 and §5.4.2).

6.1.4 Overall System Response due to a Particle in the Transducer

The equations which describe the component parts of the system have been collected in the above sub-sections and may be combined to yield overall equations which describe the performance of the system as a whole. Particles passing through the coil transducer generate pulses at the outputs of both demodulators and we write the pulse heights as V_{FMo} and V_{AMo}.

For FM perturbations, combining Equations 6.1.3(1), 6.1.2(6) and 6.1.1(1) produces

$$V_{FMo} = G_{FM}\left(-\frac{1}{2}\right) f K_{co} \mathrm{Re}[D]$$

$$= K_{FM}\,\mathrm{Re}[D] \tag{1}$$

where K_{FM} is defined as the FM system constant.

For AM perturbations, combining Equations 6.1.3(2), 6.1.2(3) and (5), and 6.1.1(2) yields

$$V_{AMo} = G_{AM} V_o Q \left(\frac{G}{V_o} \cdot \frac{dV_o}{dG}\right) K_{co} \mathrm{Im}[D]$$

$$= K_{AM}\,\mathrm{Im}[D] \tag{2}$$

where K_{AM} is defined as the AM system constant.

These equations indicate that given the complex D value of a particle, and having obtained the values of the overall system constants K_{FM} and K_{AM}, we can predict the heights of the system output pulses generated by the particle passing through the transducer.

6.2 SIGNAL TO NOISE RATIO CALCULATIONS FOR THE SYSTEM

In Chapter 4 we analysed the effects of noise on the frequency and amplitude of the marginal oscillator signal. In this section we develop expressions for the signal to noise ratios of the system and the effects of changing tuned circuit parameters.

6.2.1 The Derivation of the Signal to Noise Ratios

Rewriting Equation 4.2.4(3):

$$f_{n\,rms} = \frac{1}{4\pi}\left(\frac{1}{2^{1/2}CV_o}\right) I_{n\,rms} \tag{1}$$

which is the rms frequency noise on the marginal oscillator signal.

Similarly, rewriting Equation 4.2.2(3)

$$V_{n\,rms} = \frac{1}{G}\left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right)\frac{I_{n\,rms}}{2^{1/2}} \tag{2}$$

which is the rms noise voltage of the amplitude envelope of the marginal oscillator signal.

We define a system signal to noise ratio (SNR) for both FM and AM effects to be the ratio of the signal voltage to the rms noise voltage at the system output, and we must perform the analysis on both the FM and AM demodulator output signals. Thus, combining Equation 6.1.4(1) and Equation (1) above yields

$$SNR_{FM} = \frac{V_{FMo}}{G_{FM}f_{n\,rms}} = \frac{G_{FM}\left(-\frac{1}{2}\right)fK_{co}Re[D]}{G_{FM}\cdot\frac{1}{4\pi}\left(\frac{1}{2^{1/2}CV_o}\right) I_{n\,rms}} \tag{3}$$

which simplifies to

$$SNR_{FM} = \left[\frac{2^{1/2}V_o\cdot(C/L)^{1/2}K_{co}}{I_{n\,rms}}\right] Re[D] \tag{4}$$

Similarly, combining Equation 6.1.4(2) and Equation (2) yields

$$SNR_{AM} = \frac{V_{AMo}}{G_{AM}V_{n\,rms}} = \frac{G_{AM}V_o\frac{\omega C}{G}\left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right)K_{co}Im[D]}{G_{AM}\cdot\frac{1}{2^{1/2}}\cdot\frac{1}{G}\left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right) I_{n\,rms}} \tag{5}$$

which simplifies to

$$SNR_{AM} = \left[\frac{2^{1/2}V_o.(C/L)^{1/2}K_{co}}{I_{n\ rms}}\right] Im[D] \qquad (6)$$

6.2.2 Signal to Noise Ratio Equations: Implications for System Operation

Having derived expressions for the system signal to noise ratios, it is now appropriate to investigate the effects of parameter variations. First, we identify key features of the signal to noise ratio equations.

- The signal to noise ratio for both FM and AM effects are linked to the Re[D] and Im[D] components via the same (square bracketed) expression.
- The SNR varies inversely with the rms noise current level, $I_{n\ rms}$
- The SNR is proportional to the coil K_{co} value.
- The SNR is proportional to the product $V_o.(C/L)$, which must be treated as a single term since V_o is, in general, a nonlinear function of (C/L).

Thus, to maximise the SNR's, we would wish to maximise the values of Re[D] and Im[D], K_{co} and $V_o.(C/L)$ whilst minimising the value of $I_{n\ rms}$.

The values of Re[D] and Im[D] may be maximised for a given particle size by working at a frequency such that the value of the a/δ ratio produces significant values of both $Re[D_n]$ and $Im[D_n]$ (see Figure 3.2.1(b)).

The effect of rms noise current (see §4.2) may be reduced by low-pass filtering in the demodulation stages but such filtering must not be so severe as to significantly distort the signal pulses.

A high value of K_{co} (see §3.4) may be obtained by minimising coil dimensions or by arranging for the particle to pass close to the coil windings.

The expression $V_o.(C/L)$ can be simplified for the Robinson oscillator configuration but for the van der Pol oscillators, V_o is a nonlinear function of (C/L) and this complicates matters. For a given coil transducer, the values of the coil inductance, L, and series resistance, R, are fixed and changing the system operating frequency involves changing the tuned circuit capacitor, C. Say we wish to increase the frequency, we would have to reduce the value of C which would, in turn, reduce the value of the term $(C/L)^{1/2}$. The value of V_o would, however, increase because the value of R is multiplied by the factor (C/L) to yield G, the equivalent parallel tuned circuit conductance which obviously reduces. For the Robinson oscillator, where the tuned circuit is fed with an AC current source of fixed peak magnitude I_o, the peak

voltage signal is given by $V_o=(I_o/R)(L/C)^{1/2}$. For the van der Pol oscillator, the nonlinear amplifier characteristic causes the term $V_o.(C/L)^{1/2}$ to be proportional to $(L/C)^x$. where $x \geq 1/2$ and depends on the nonlinearity of the oscillator amplifier I_o-V_o characteristic.

6.3 ESTIMATED VALUES OF SUB-SYSTEM AND SYSTEM KEY PARAMETERS

In this section we obtain estimates of the values of key sub-system parameters by considering transducer geometry and circuit components.

6.3.1 The Inductive Coil Transducer: Theoretical Estimates of Key Parameters

A 26-turn inductive coil transducer with dimensions, length $\lambda=6$mm, radius $A=3$mm, wound from 35 SWG enamelled copper wire (diameter, $d\simeq 212\mu$m) can be described in terms of its inductance, L, its series resistance, R, and a geometric coil constant, K_{co}. The value of L may be approximated by Equation 3.4.1(6):

$$L = K_n \left[\mu_o n^2 \pi A^2 \lambda\right] \tag{1}$$

where

$$K_n = \frac{1}{1 + 0.9(A/\lambda) - 0.02(A/\lambda)^2} \tag{2}$$

Inserting the values yields $L\simeq 2.8\mu$H.

The value of R is given by Equation 3.4.4(1):

$$R = \frac{1}{2} R_{dc} \cdot \left(\frac{d}{\delta}\right).\Phi \tag{3}$$

where, for operation at 10MHz, $\delta=20.9\mu$m and $d/\delta=10.2$. Also, using Figure 3.4.4(a), the value of $\Phi\simeq 3.2$ and noting winding geometry and the resistivity of copper ($\sim 1.7\times 10^{-8}$ Ωm) yields the value $R_{dc}\simeq 0.223\Omega$. Inserting these values into Equation (3) yields a value of $\sim 1.9\Omega$ for R.

From §3.4.1

$$K_{co} = K_c(0,0) = \left[\frac{2\sin^2\alpha_o}{K_n \lambda A^2}\right] \tag{4}$$

where $\alpha=45°$ in this case, and substituting values yields a value $K_{co}=26.8\times 10^6$ m^{-3}.

6.3.2 The Marginal Oscillator: Theoretical Estimates of Key Parameters

Owing to the difficulty in developing an analytic solution to the Marginal Oscillator parameters, we resorted to using a computer model to obtain estimates of the key parameters V_o and $(G/V_o.dV_d/dG)$. The basic marginal oscillator configuration used for both experimental work and the computer simulations is shown in Figure 6.3.2(a).

The oscillator circuit was simulated using the HPSPICE programme. As with a practical circuit, an initial perturbation must be applied to start oscillations and we therefore arranged for a 100ns, 100μA pulse to be applied to the one of the nodes. The oscillator signal builds up, as shown in Figure 6.3.2(b), and after about 6μs, stabilises at the levels shown. We have also plotted, in the Figure, the steady-state waveforms at the same nodes between 20μs and 25μs. Over this period we performed a Fourier transform on the waveforms to obtain values of the fundamental components. We define the voltage at node 6 to be the point at which we derive the oscillator voltage V_o. We obtain from the simulation a value $V_o \simeq 5.45V$.

The required value of the factor $(G/V_o.dV_d/dG)$ is obtained by first considering the value of conductance, G*, seen by the effective current-source output of JFET2, the value of this conductance being given approximately by:

$$G^* = G + 1/RD2 \qquad (1)$$

where G is the tuned circuit conductance and RD2 is the drain resistor of the second JFET.

A change in G must be translated to a change ΔG^* before an approximation to the effective value of $(G/V_o.dV_d/dG)$, which we write as $(G^*/V_o.dV_d/dG^*)$, can be obtained. The value of G, and hence G*, was changed by a small amount and the corresponding value of V_o obtained by running the new values in the simulation programme. We thus obtained a value of -3 for the factor $(G^*/V_o.dV_d/dG^*)$.

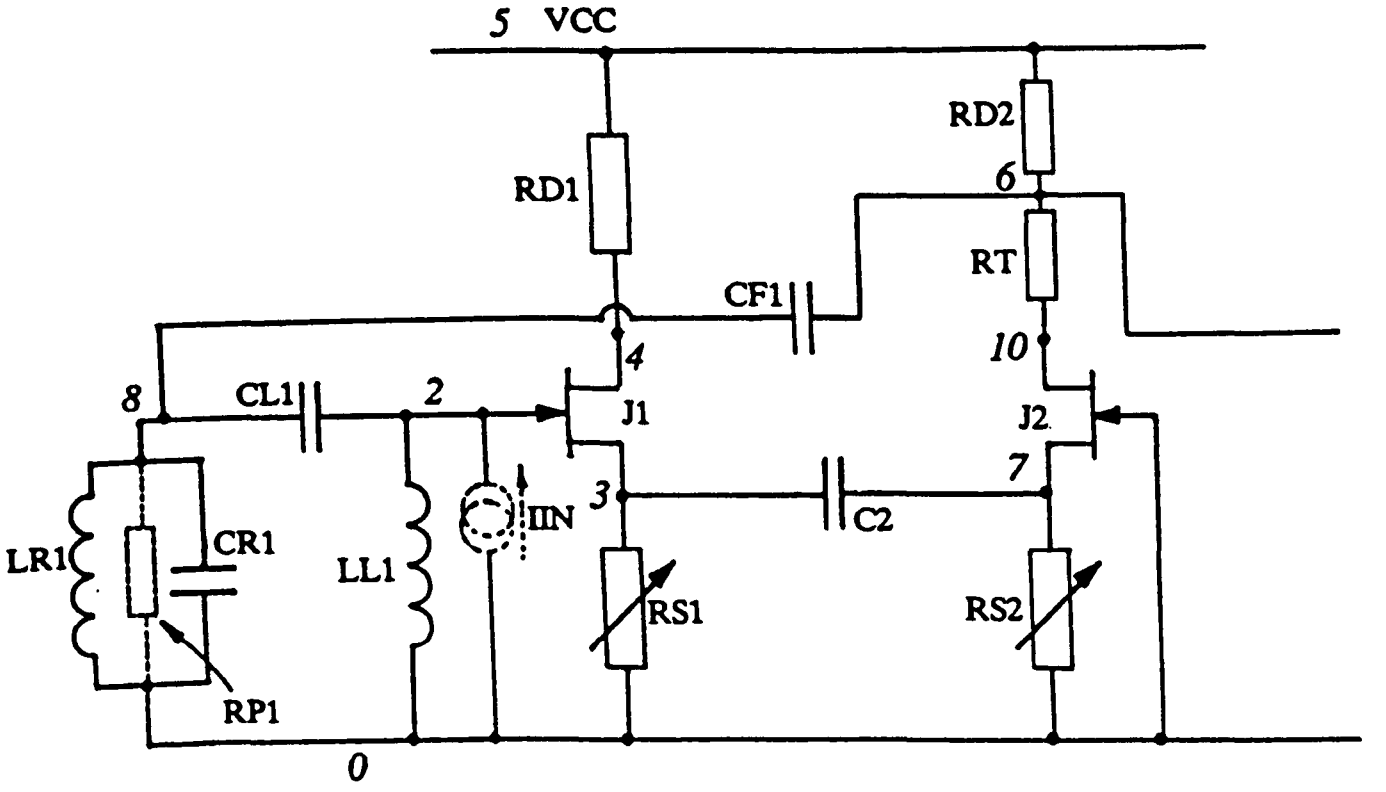
6.3.3 FM Demodulator: Theoretical Estimates of Key Parameters

Referring to Figure 5.2.1(a) we note that the FM demodulator may be divided into two stages: the pulse stage (whose average value represents the demodulated signal) and the subsequent active filter stage (whose output *is* the demodulated signal).

The average value of the output signal from the pulse stage is given by

$$V_{AVE} = K_{PS} f_{IN} \qquad (1)$$

where

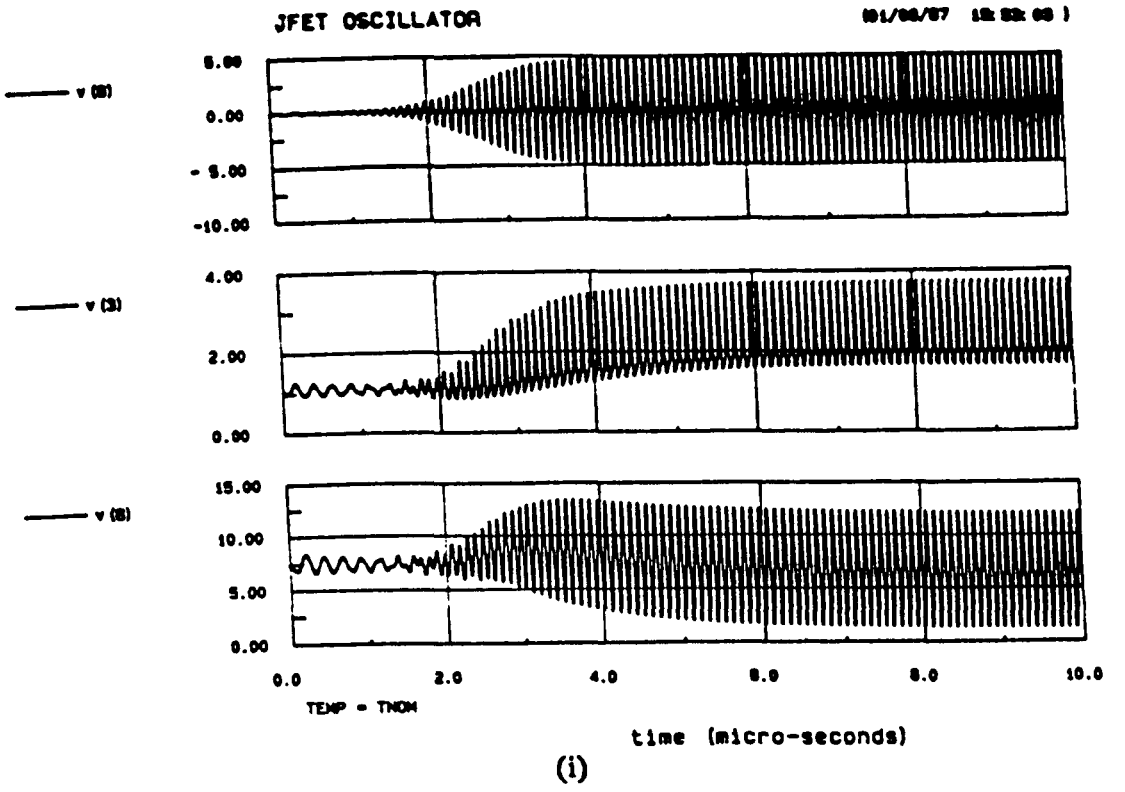


(i)

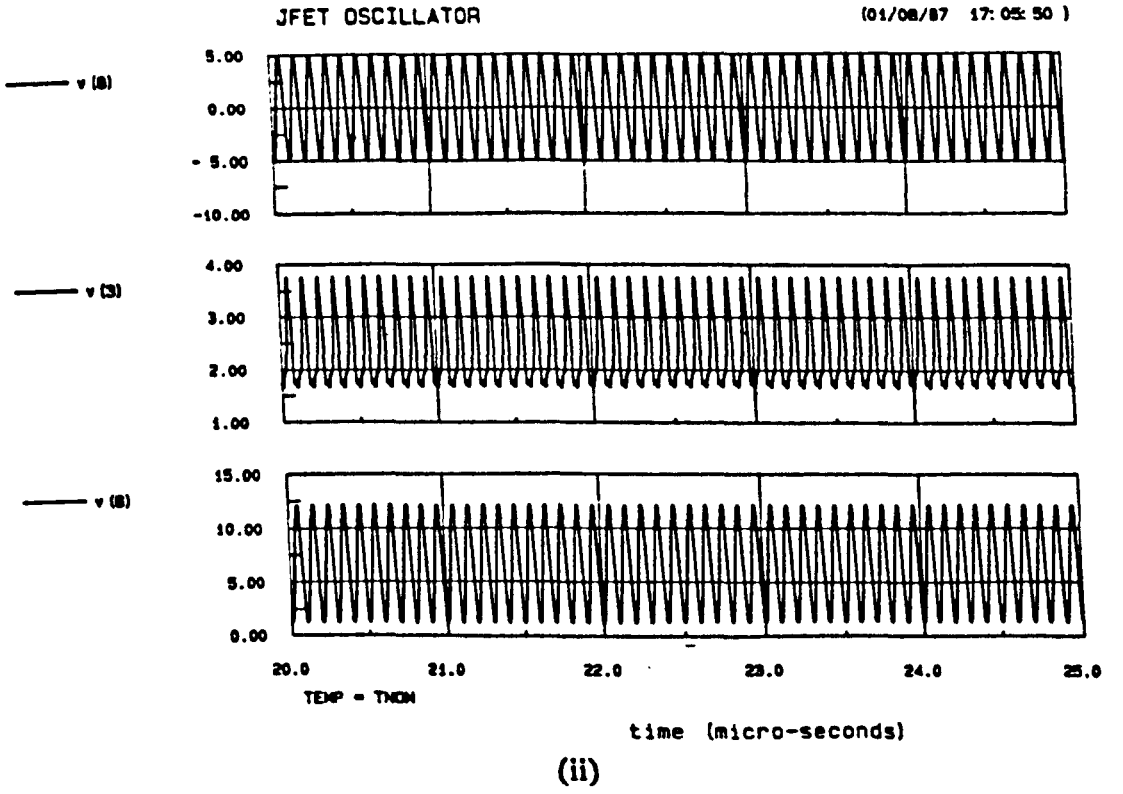
```
****** 03/18/87 ********  HPSPICE  (2.0  2330)  ******* 12:20 34 ******

 JFET OSCILLATOR

****     INPUT LISTING                    TEMPERATURE =   27.000 DEG C

*******************************************************************************


 RS1 = 3 0 1.08K
 RD1 = 5 4 2.2K
 J1 = 4 2 3 JMOD1
 C2 = 3 7 1500P
 RS2 = 7 0 737
 RD2 = 5 6 6.8K
 J2 = 10 0 7 JMOD2
 RT = 10 6 1
 CF1 = 6 8 12P
 RP1 = 8 0 16.3K
 CR1 = 8 0 90.465342P
 LR1 = 8 0 2.8U
 LL1 = 2 0 100U
 CL1 = 8 2 10P
 VCC = 5 0 DC 17
 IIN 2 0 PULSE(0 100U 0NS .25NS .25NS 100NS)
 .MODEL JMOD1 NJF BETA=1.18M VTO=-2.0 CGS=4.5P CGD=1.2P
 .MODEL JMOD2 NJF BETA=0.764M VTO=-2.4 CGS=4.5P CGD=1.2P
 .TRAN 1NS 25000N 20000N 1N
 .OPTIONS ITL5=10000000 LIMPTS=2000000
 .GRAPH TRAN V(2) V(3) V(4) V(6) V(7) V(8) V(10,6)
 .FOUR 10MEG V(2) V(3) V(4) V(6) V(7) V(8) V(10,6)
 .END
```

(ii)

Figure 6.3.2(a) The HPSPICE simulation of the marginal oscillator used in this work; (i) the circuit and (ii) the corresponding input listing.



(i)



(ii)

Figure 6.3.2(b) The transient waveforms from the HPSPICE simulation; (i) from start-up, between 0μs and 10μs and (ii) after the steady-state has been reached, between 20μs and 25μs.

$$K_{PS} = (V^+ - V^-)\tau \qquad (2)$$

and where the pulse output levels, $V^+ = -V^- = 8.8V$, and the period of the monostable in the circuit, τ, is 7.95μs. The corresponding value of K_{PS} is therefore approximately 140μV/Hz.

The output signal from the pulse stage is fed into the active low-pass filter which, at frequencies well below the 3dB cut-off frequency, produces a gain of approximately 8.

By multiplying the pulse stage gain with that of the subsequent filter, we obtain an estimate of the FM demodulator gain, G_{FM}, of approximately 1.1mV/Hz.

6.3.4 AM Demodulator: Theoretical Estimates of Key Parameters

The AM diode detector is preceded by an RF amplifier and is followed by an audio frequency amplifier (see Figures 5.4.1(a) and 5.4.2(a)) The gain of the AM demodulator may be estimated from the product of the amplifier gains.

The RF amplifier gain may be approximated by the gain of the bipolar transistor cascode configuration which is given by the ratio of the values of the 5.6kΩ load resistor to the value of the 1.2kΩ input resistor i.e. ~4.7.

The audio amplifier gain is due mainly to that of the operational amplifier non-inverting amplifier which has a gain given by the ratio of its resistors i.e. 510kΩ/10kΩ≃51.

The product of the above gains, ~240, is the required estimate for the magnitude of the AM demodulator gain G_{AM}. Note that due to the direction of the diode in the envelope detector, G_{AM}≃-240.

6.3.5 Overall System Constants: Theoretical Estimates

The estimated values of the sub-system key parameters, obtained in §6.3, can be combined to yield an estimate for the overall system constants, K_{FM} and K_{AM}, by using Equations 6.1.4(1) and 6.1.4(2):

$$K_{FM} = G_{FM}\left(-\frac{1}{2}\right) f K_{co} \qquad (1)$$

$$K_{AM} = G_{AM} V_o Q \left(\frac{G}{V_o}\cdot\frac{dV_o}{dG}\right) K_{co} \qquad (2)$$

It should be noted that the required value of Q must be the loaded value, Q^*≃25, and

the frequency of operation, f≃10MHz. The value of $(G/V_o.dV_o/dG)$ must also be the loaded value, $(G^*/V_o.dV_o/dG^*)\simeq -3$. Thus inserting these as well as other values obtained previously in this section yields the values $K_{FM} = -1.47\times10^{11}$ V/m^3 and $K_{AM} = 26.3\times10^{11}$ V/m^3.

6.4 SUMMARY AND CONCLUSIONS FOR CHAPTER SIX

This chapter is summarised and concluded as follows:

- Expressions for the effect on each system component (i.e. the inductive transducer, the marginal oscillator and the demodulators) of a particle passing through the transducer are combined to yield overall system equations for the demodulator output pulse heights:

$$V_{FMo} = K_{FM}\,\mathrm{Re}[D]$$

and

$$V_{AMo} = K_{AM}\,\mathrm{Im}[D]$$

the parameters K_{FM} and K_{AM} being the overall system constants for FM and AM perturbations.

- Expression for the the signal to noise ratios, defined as the ratio of the output pulse height for a channel to the rms noise voltage, have been derived. For the simple oscillator noise model used, the constant of proportionality between the Re[D] and Im[D] components and, respectively, SNR_{FM} and SNR_{AM}, is the same. Moreover, the degree of oscillator feedback amplifier nonlinearity does not appear in the constant.

- Estimated values, based on considering transducer geometry and circuit components, have been obtained for the system constants: $K_{FM} = -1.47\times10^{11}$ V/m^3 and $K_{AM} = 26.3\times10^{11}$ V/m^3

CHAPTER SEVEN

THE EXPERIMENTAL SYSTEM: CALIBRATION AND TESTING

In this chapter the experimental system and its calibration and testing are described.

7.1 EXPERIMENTAL SET-UP

In this section, we describe the the construction and mode of operation of the various component parts of the experimental set-up.

7.1.1 The Coil Transducer

The transducer used is, in effect, a single layer inductive coil which was constructed by bonding (using epoxy resin) SWG 35 enamelled copper wire onto a section of glass tubing with an outer diameter of 6mm and an inner diameter of 3mm. This is shown in Figure 7.1.1(a). The thickness of glass tubing was sufficient to exclude particles from the region close to the coil windings and, in addition, ensured that the effect of a particle on the transducer impedance could not be significantly dependent on radial position.

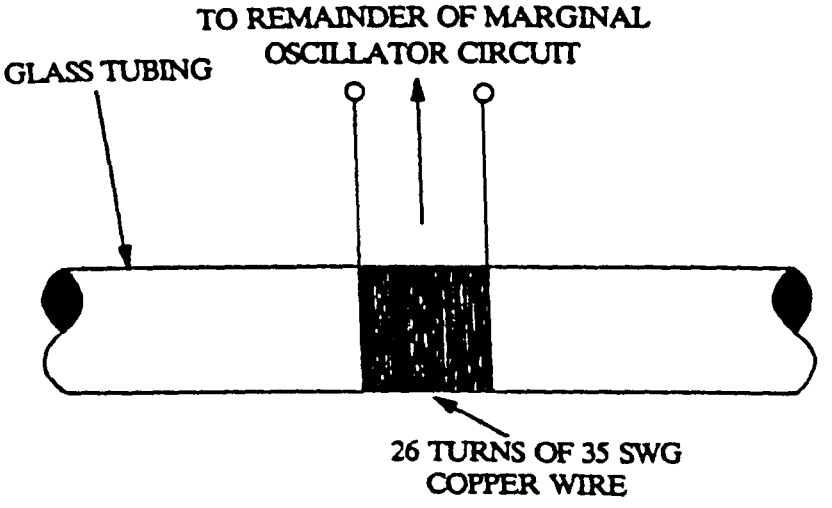
7.1.2 Marginal Oscillator Circuitry

The oscillator was constructed on a custom-made printed circuit board. Components were mounted on the patterned side of the board, the underside serving as a ground plane to which all grounded sections of the patterned side were through-connected.

The circuit used was that of Figure 4.4.2(a) with the tuned circuit capacitance due mainly to the ~50cm of coaxial cable connecting the main oscillator to the coil transducer. It was found that the addition of ~13pF was sufficient to set the frequency of oscillation at ~10.063MHz. Silvered mica capacitors and 1% metal film resistors were used throughout.

7.1.3 AM Demodulation Circuitry

The AM demodulation process was performed at the marginal oscillator frequency (~10MHz) and the printed circuit board techniques described in §7.1.2 were also used for the circuitry.



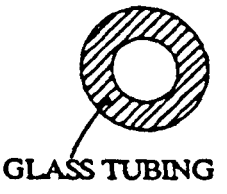


Figure 7.1.1(a) The inductive coil transducer; (i) side view and (ii) cross-sectional end view.

7.1.4 FM Demodulator

The FM demodulator was constructed on two Vero 'microboard' ground-plane stripboards. Use of printed circuit techniques did not seem justified because, after initial downconversion, the effective carrier frequency was only ~63kHz.

7.1.5 Sample Rods and the Associated Mechanical Reciprocating Mechanism

The particles used to calibrate and test the system were bonded to plastic sample rods which were, in turn, inserted into a mechanical reciprocating mechanism (see Figure 7.1.5(a)). By varying the voltage applied to the electric motor, the sample rod, and hence the particle, was passed through the transducer at different speeds. For each sample, the particle was positioned at a point on the rod which ensured that the particle passed through the transducer at the maximum possible speed.

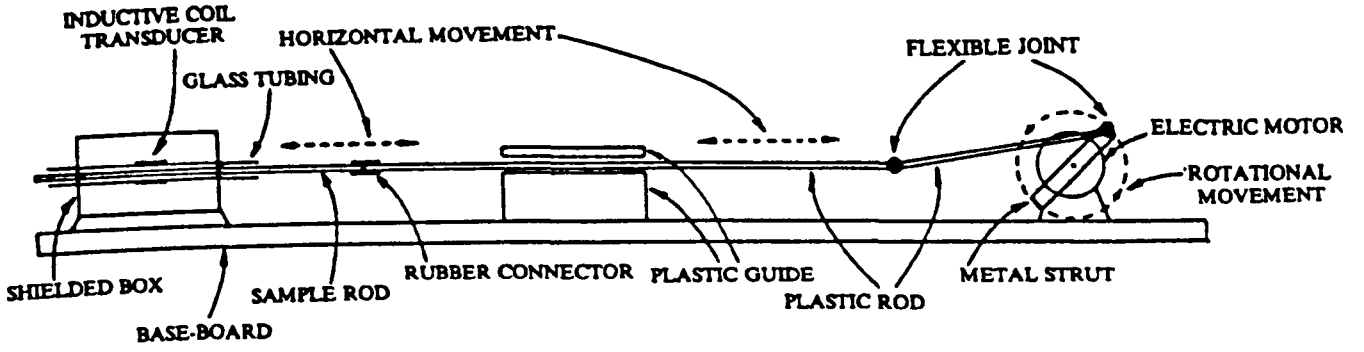
It was found experimentally that the speed of the particle through the coil could be varied in the range 0.4m/s to 4.0m/s. The minimum speed was limited by frictional effects whilst the maximum speed was limited mainly by mechanical resonances. Such speeds are typical of the range of particle velocities likely to be met in full scale lubrication systems.

7.2 SYSTEM CALIBRATION

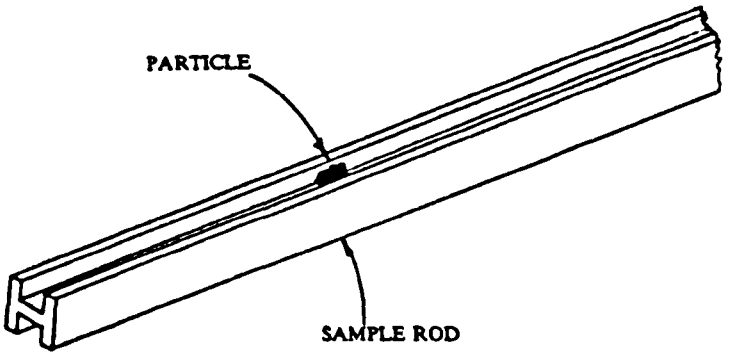
Preliminary tests on the system using metal particles of various materials indicated qualitative agreement with the theoretical effects discussed in previous chapters. However, the values for the overall system constants, K_{FM} and K_{AM}, were found experimentally to be lower than the values obtained from the key parameter estimates of §6.3. We therefore calibrated the system using particles of a known material with approximately known sizes. Particles of lead were used for this purpose.

7.2.1 Calibration Using Lead Particles

Values of V_{FMo} and V_{AMo}, obtained at the system outputs on passing a number of different lead particles through the coil transducer using the mechanism described in §7.1.5, were plotted as shown in Figure 7.3.2(a). We found that using the values $K_{FM} = -1.18\times10^{11}$ V/m^3 and $K_{AM} = 5.9\times10^{11}$ V/m^3 produced a good fit to the practical points and the corresponding theoretical curve for the particle diameter range 100-400μm has been added to the Figure.



(i)



(ii)

Figure 7.1.5(a) (i) A schematic diagram of the mechanical reciprocating mechanism used to calibrate and test the system, and (ii) a schematic diagram of a sample rod with a particle affixed.

7.2.2 System Noise Level Measurements

We found that the noise signals at the outputs of the demodulator circuits could be attributed to two sources:

(i) Electronic noise

(ii) Microphonic (vibrational) noise

While the mechanical reciprocating mechanism was not being operated, the microphonic noise was insignificant. By measuring the system outputs while the mechanical system was dormant, we obtained the noise waveforms shown in Figure 7.2.2(a)(i). The peak-to-peak values of the output noise voltages were approximately 10mV and 80mV for the FM and AM channels; the corresponding values of the rms noise voltages being $V_{FM\ n\ rms} \simeq 1.7mV$ and $V_{AM\ n\ rms} \simeq 13mV$.

With the mechanical reciprocating mechanism operating, microphonic noise augmented the electronic noise. Using a sample rod *without* a sample attached, we obtained noise waveforms as shown in Figure 7.2.2(a)(ii). The approximate peak-to-peak values of the output noise voltages were 30mV and 100mV for the FM and AM channels; the corresponding values of the rms noise voltages being $V_{FM\ n\ rms} \simeq 5mV$ and $V_{AM\ n\ rms} \simeq 17mV$.

7.3 SYSTEM TESTS USING PARTICLES

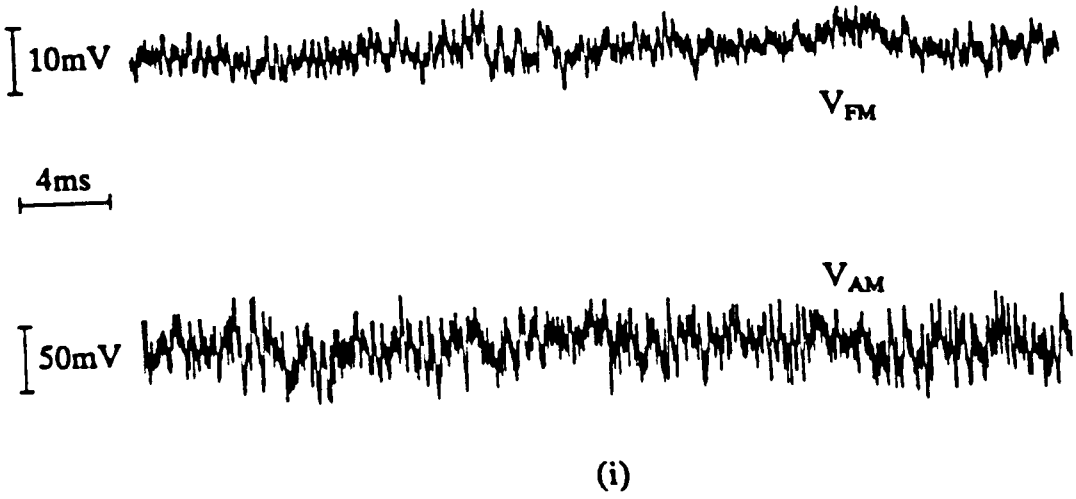
Having calibrated the system using lead particles it was necessary to test the system, both for effects of particle speed variation and for different materials.

7.3.1 System Output Characteristics for Different Speeds

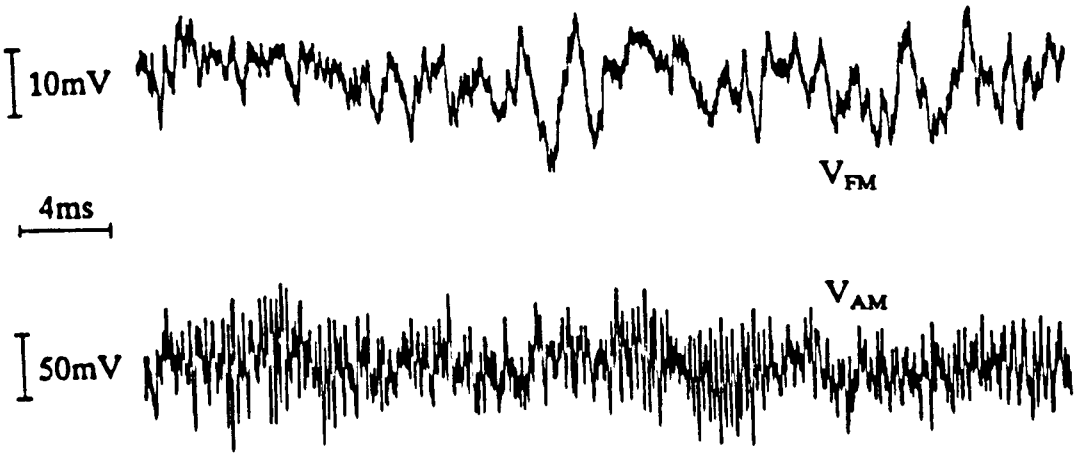
The mechanical reciprocating mechanism allows variations in particle speed within the approximate range 0.4m/s and 4.0m/s (see §7.1.5). We tested the effect of speed variations by arranging to pass the same (copper) particle through the transducer at four different speeds: approximately 0.4m/s, 1.4m/s, 2.9m/s and 4.0m/s. The resultant pulses are shown in Figure 7.3.1(a). The key point to note is that there is no measurable variation in pulse height over the speed range.

7.3.2 System Characteristics for Particles of Different Materials

Sample rods, each with a particle attached, were prepared for seven materials. In each case particles were cut or broken from solid pieces of the material and no attempt was made to select specific shapes although particles in the approximate size range 100-



(i)



(ii)

Figure 7.2.2(a) System noise outputs (i) whilst the mechanical reciprocating mechanism is *not* in operation, and (ii) whilst the mechanism is operated using a sample rod without a particle attached.

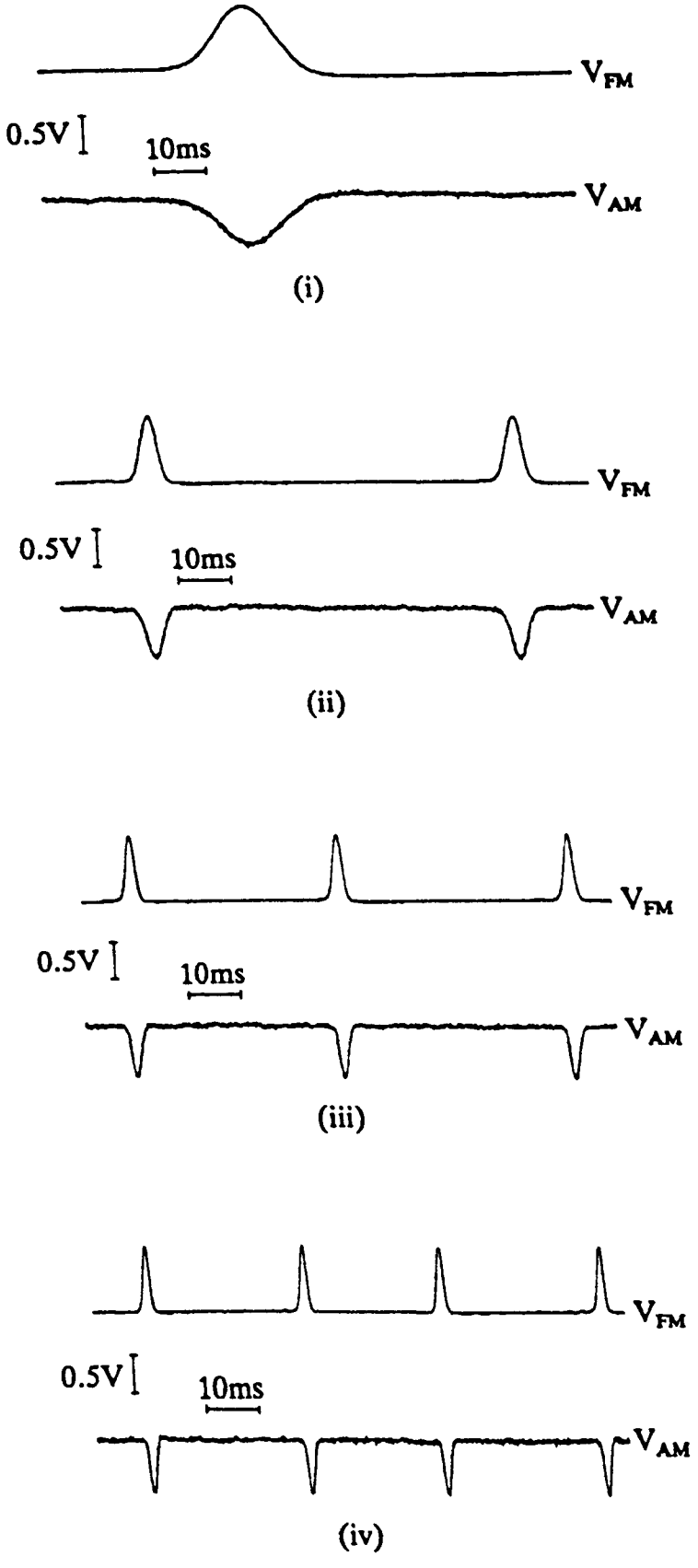


Figure 7.3.1(a) System output pulses due to a copper particle being passed through the transducer at approximate speeds; (i) 0.4m/s, (ii) 1.4m/s, (iii) 2.9m/s, and (iv) 4.0m/s.

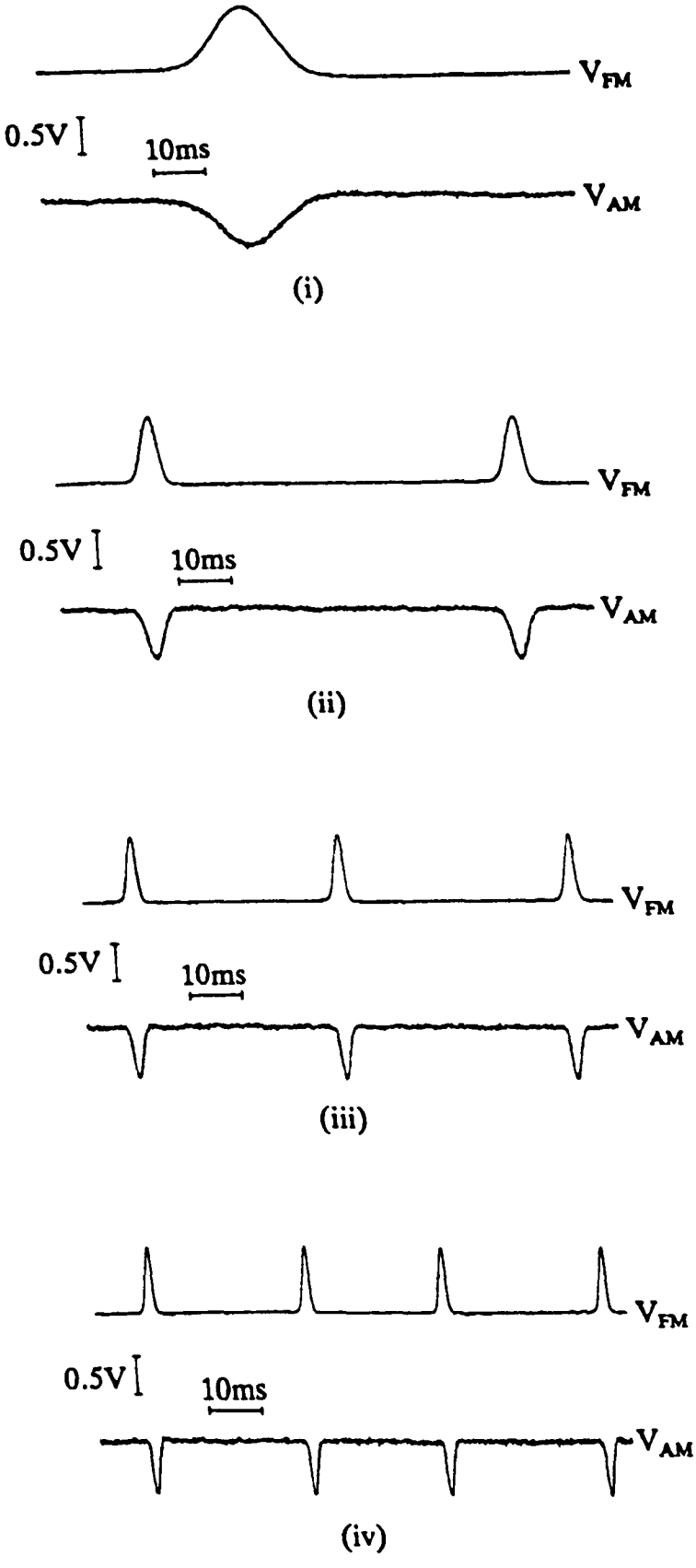


Figure 7.3.1(a) System output pulses due to a copper particle being passed through the transducer at approximate speeds; (i) 0.4m/s, (ii) 1.4m/s, (iii) 2.9m/s, and (iv) 4.0m/s.

400μm were chosen. This size range corresponds to particles worthy of monitoring within a lubrication system.

The materials used were:

(a) lead ($\rho=21\times10^{-8}\Omega$m, $\mu_r=1$, @ 20°C) [1]

(b) brass ($\rho=6\times10^{-8}\Omega$m, $\mu_r=1$, @ 20°C) [1]

(c) copper ($\rho=1.7\times10^{-8}\Omega$m, $\mu_r=1$, @ 20°C) [1]

(d) steel EN9 ($\rho\simeq18\times10^{-8}\Omega$m, $\mu_r\simeq100$, @20°C) [2]

(e) steel EN5A ($\rho\simeq18.7\times10^{-8}\Omega$m, $\mu_r\simeq800$, @20°C) [2]

(f) mu-metal (values of ρ and μ_r unavailable)

(g) ferrite B10 (4C6) ($\rho=5\times10^{2}\Omega$m, $\mu_r\simeq34$) [3,4]

V_{FMo}-V_{AMo} locus plots for each material were then made, each point representing an individual sample, and these are plotted in Figures 7.3.2(a)-(g). Also plotted in the Figures are theoretical curves for each material over the size range 100-400μm assuming system constants values $K_{FM}=-1.18\times10^{11}$ V/m^3 and $K_{AM}=5.9\times10^{11}$ V/m^3 which were obtained by calibrating using lead particles (see §7.2.1).

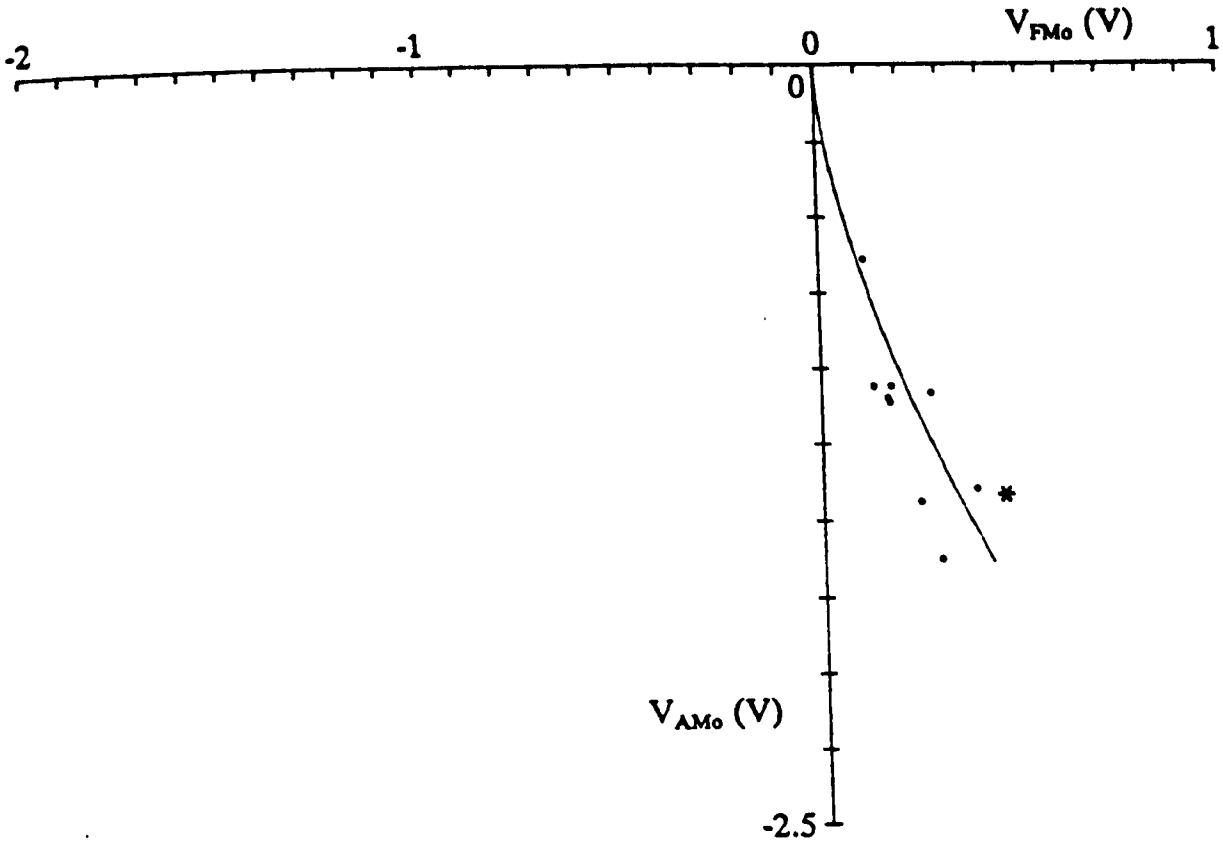
7.4 SUMMARY AND CONCLUSIONS FOR CHAPTER SEVEN

This chapter is summarised and concluded as follows:

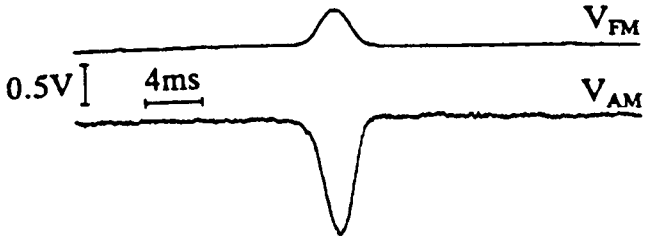
- We have described the experimental set-up which includes the circuitry described in previous chapters as well as a mechanical reciprocating mechanism used to pass particles (which are attached to plastic sample rods) through the system transducer repeatedly.
- Constructional details for the inductive transducer, the marginal oscillator circuitry and the FM and AM demodulator circuitry have been given.
- The system was calibrated using lead particles and we found that system constant values, $K_{FM}=-1.18\times10^{11}$ V/m^3 and $K_{AM}=5.9\times10^{11}$ V/m^3, produced a good fit to the practical points.
- The system output noise levels were found to be

(i) $V_{FM\,n\,rms}\simeq1.7$mV and $V_{AM\,n\,rms}\simeq13$mV, whilst the mechanical system was *not* in operation.

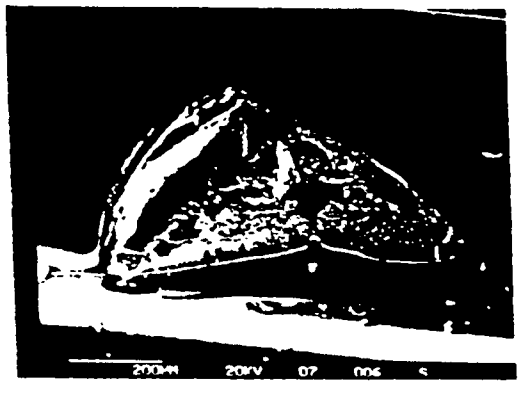
(ii) $V_{FM\,n\,rms}\simeq5$mV and $V_{AM\,n\,rms}\simeq17$mV, whilst the mechanical system *was* in operation.



(i)

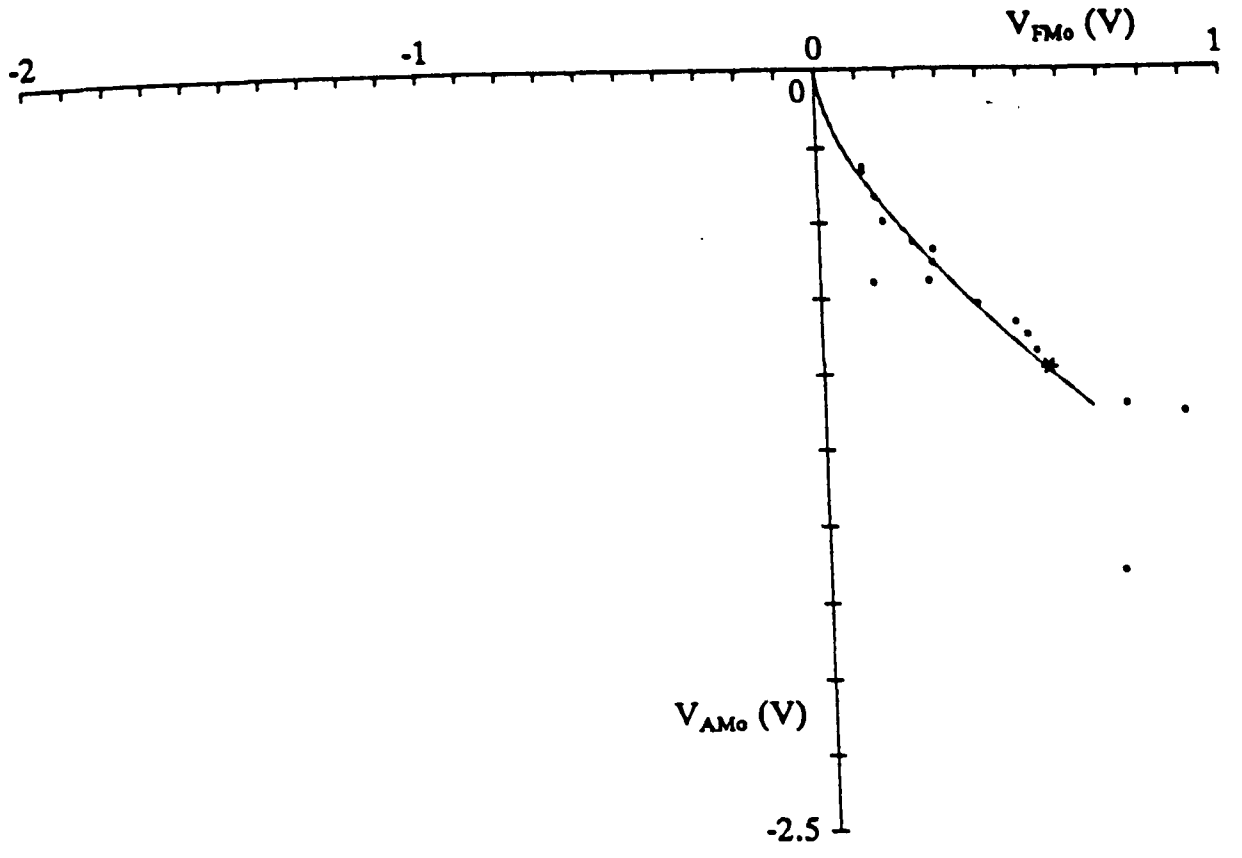


(ii)

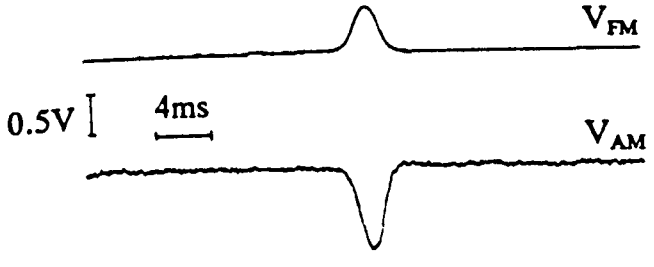


(iii)

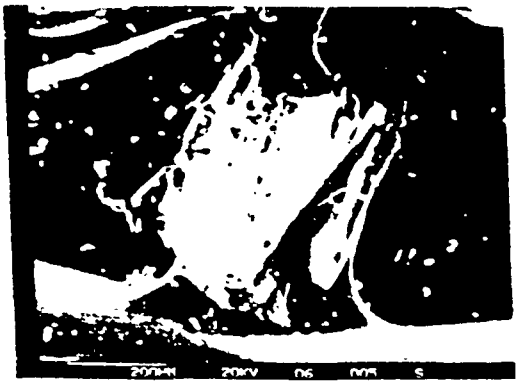
Figure 7.3.2(a) System response to lead particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (the solid curve represents the theoretical locus for particles in the size range 100μm to 400μm assuming overall system constant values, $K_{AM}=5.9\times10^{11}$ V/m^3 and $K_{FM}=-1.18\times10^{11}$ V/m^3), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.



(i)

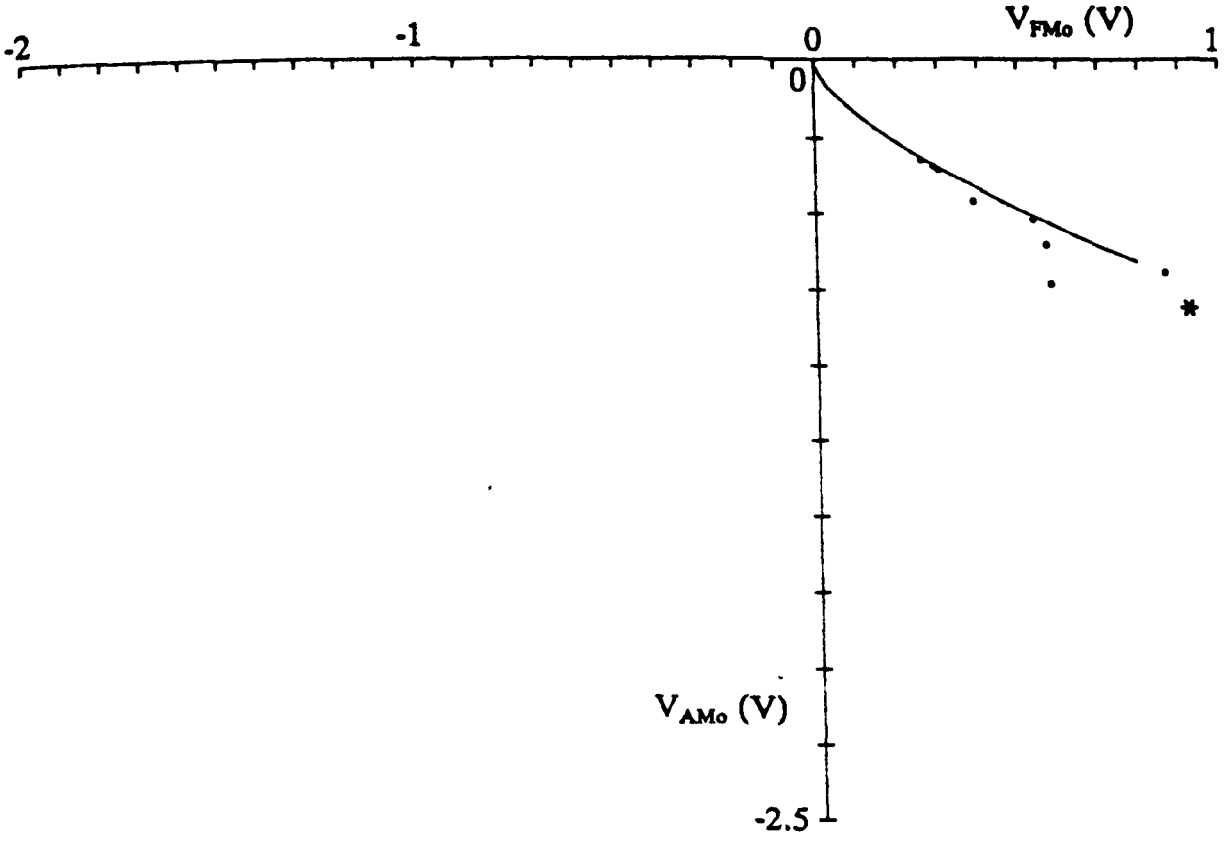


(ii)

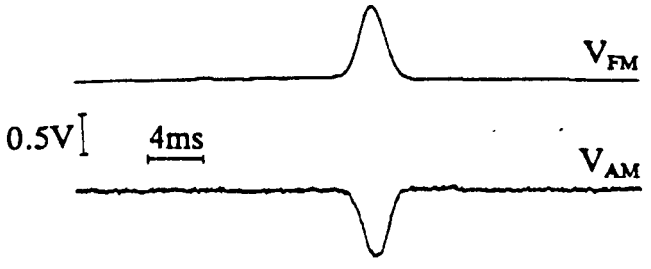


(iii)

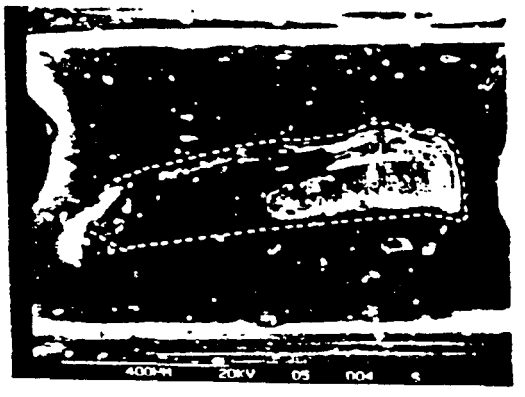
Figure 7.3.2(b) System response to brass particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (the solid curve represents the theoretical locus for particles in the size range 100μm to 400μm assuming overall system constant values, $K_{AM}=5.9\times10^{11}$ V/m^3 and $K_{FM}=-1.18\times10^{11}$ V/m^3), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.



(i)

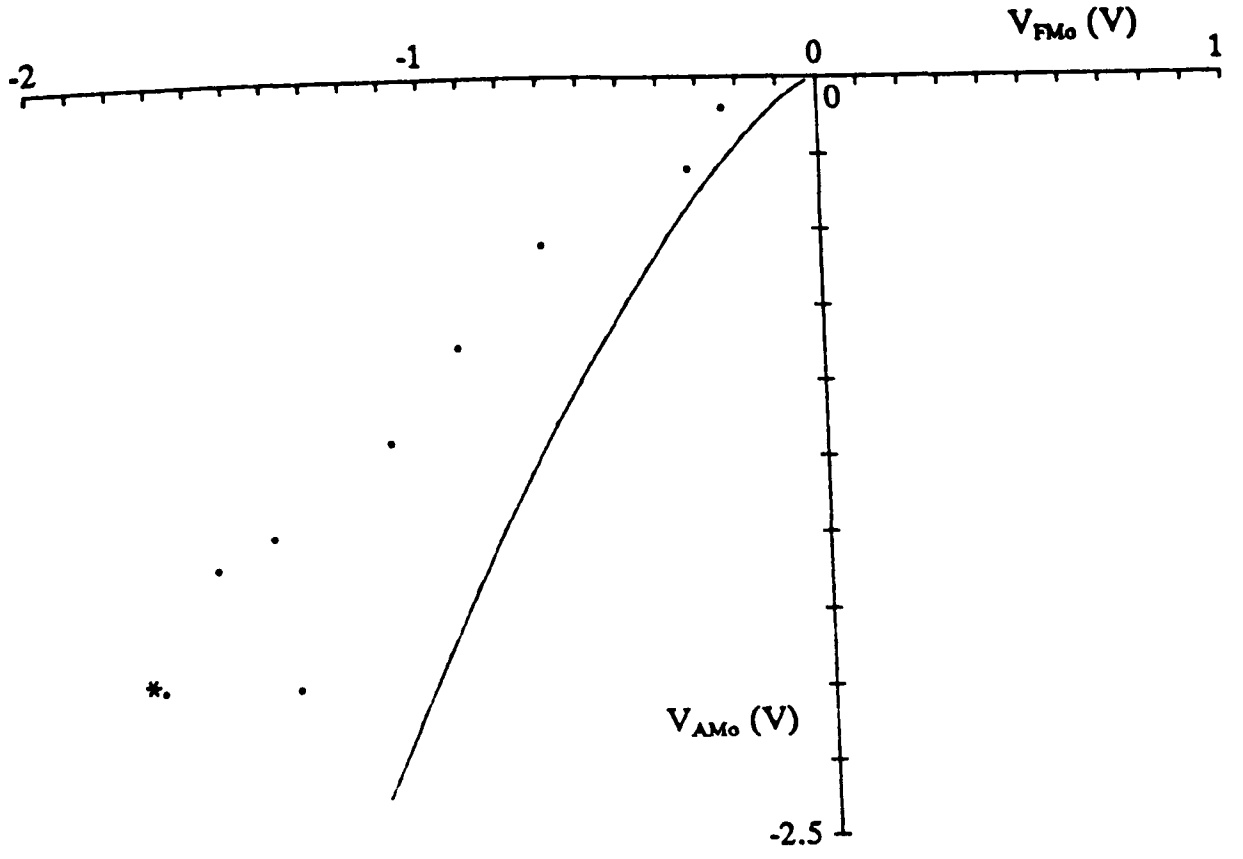


(ii)

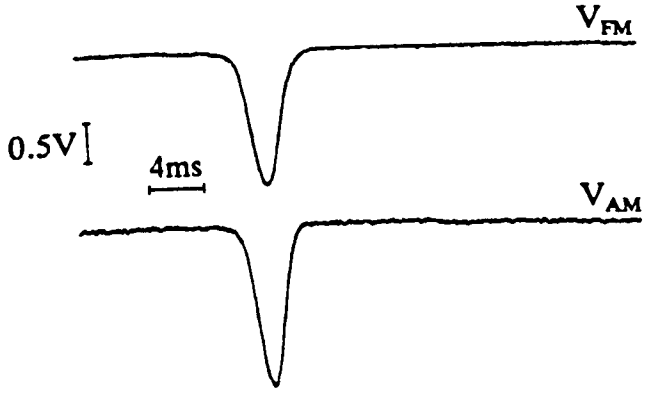


(iii)

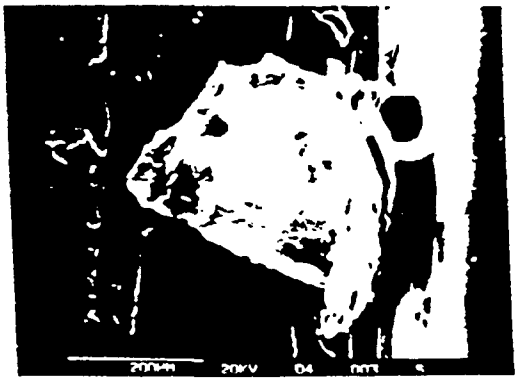
Figure 7.3.2(c) System response to copper particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (the solid curve represents the theoretical locus for particles in the size range 100μm to 400μm assuming overall system constant values, $K_{AM}=5.9\times10^{11}$ V/m^3 and $K_{FM}=-1.18\times10^{11}$ V/m^3), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.



(i)

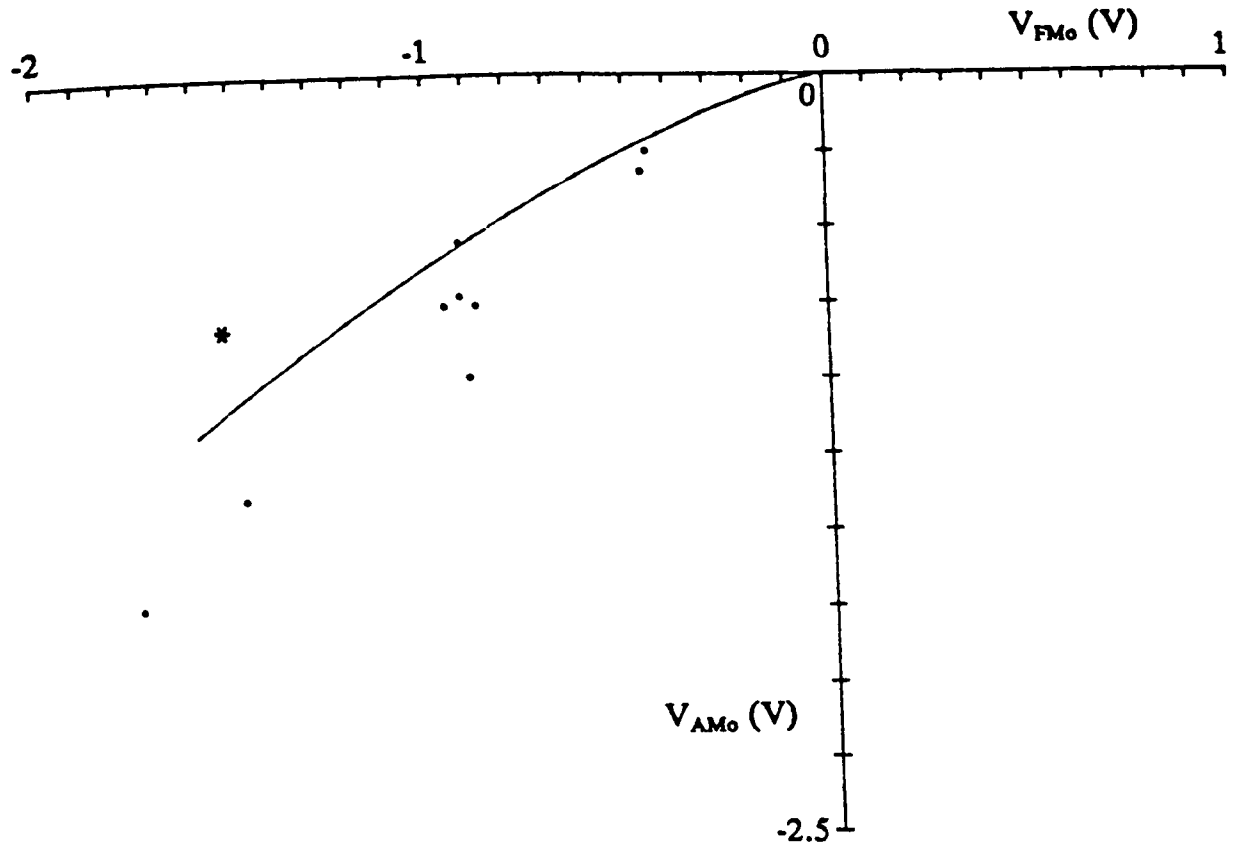


(ii)

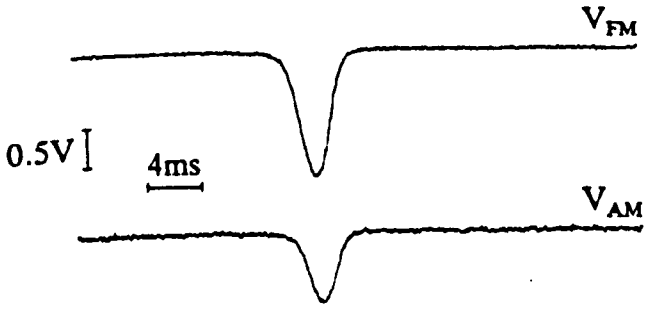


(iii)

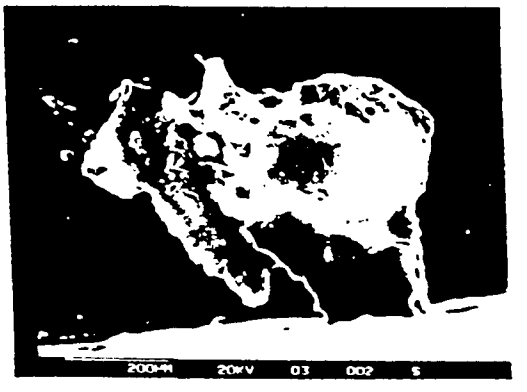
Figure 7.3.2(d) System response to steel EN9 particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (the solid curve represents the theoretical locus for particles in the size range 100μm to 400μm assuming overall system constant values, $K_{AM}=5.9\times10^{11}$ V/m^3 and $K_{FM}=-1.18\times10^{11}$ V/m^3), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.



(i)

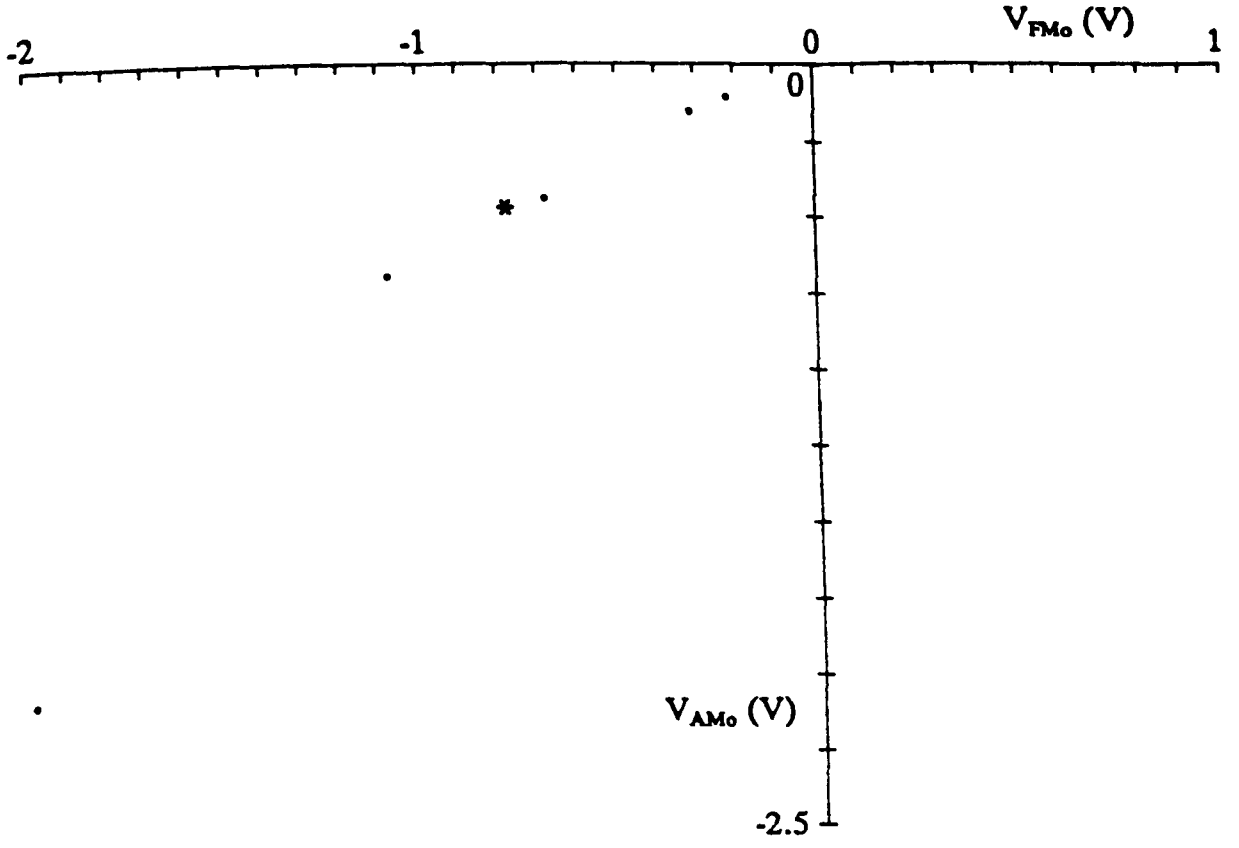


(ii)

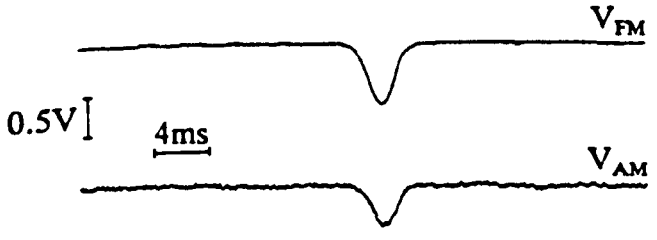


(iii)

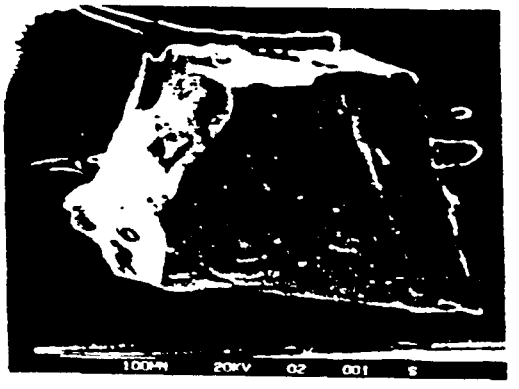
Figure 7.3.2(e) System response to steel EN5A particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (the solid curve represents the theoretical locus for particles in the size range 100μm to 400μm assuming overall system constant values, $K_{AM}=5.9\times10^{11}$ V/m^3 and $K_{FM}=-1.18\times10^{11}$ V/m^3), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.



(i)

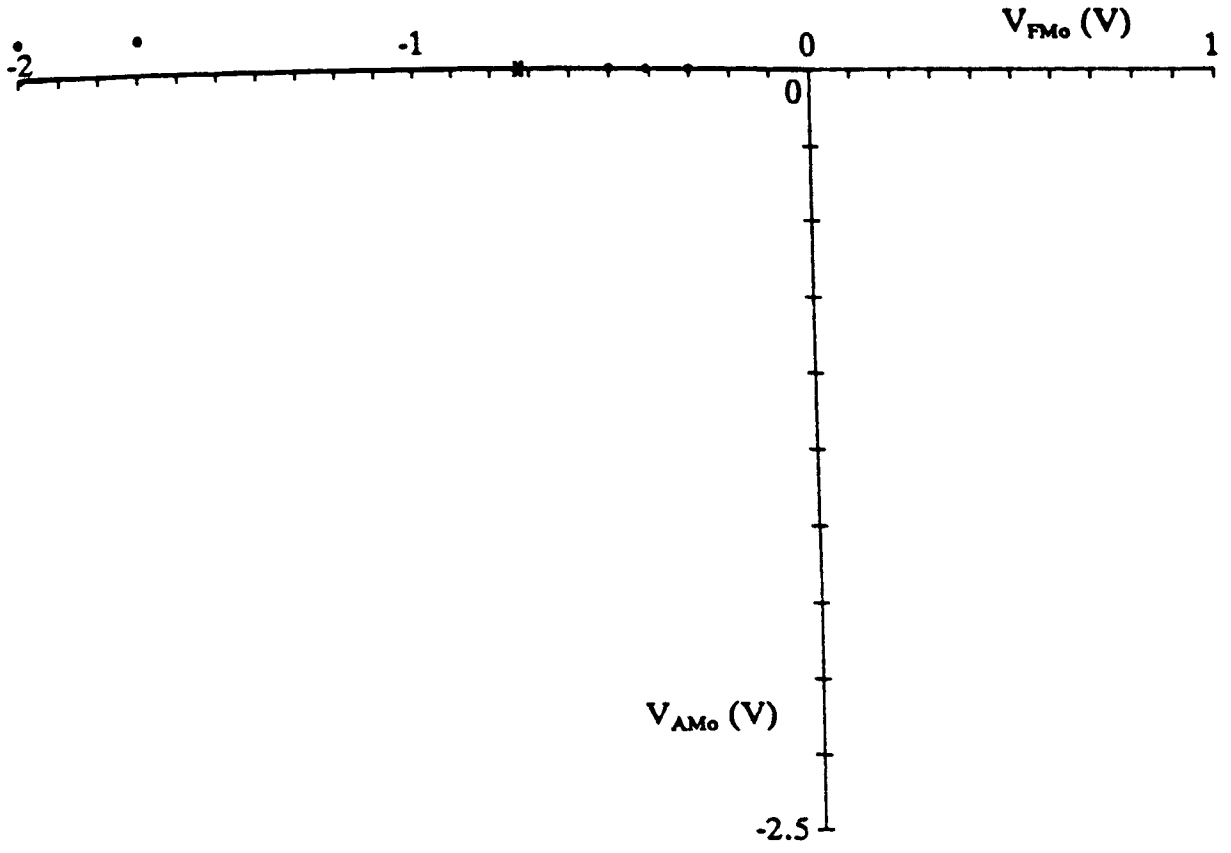


(ii)

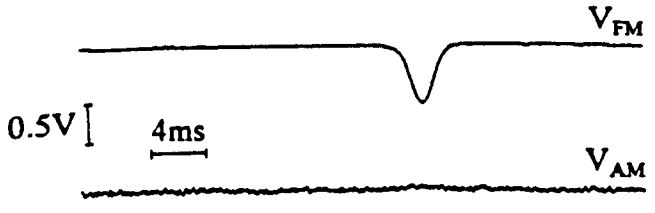


(iii)

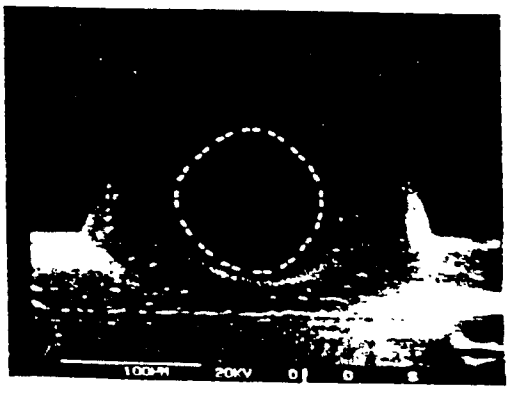
Figure 7.3.2(f) System response to mu-metal particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (a theoretical curve has not been given because of lack of data on mu-metal resistivity and magnetic permeability), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.



(i)



(ii)



(iii)

Figure 7.3.2(g) System response to ferrite particles; (i) experimental pulse heights plotted on the V_{FMo}-V_{AMo} plane (the solid curve represents the theoretical locus for particles in the size range 100μm to 400μm assuming overall system constant values, $K_{AM}=5.9\times10^{11}$ V/m^3 and $K_{FM}=-1.18\times10^{11}$ V/m^3), (ii) an example of the pulses produced by one particle (starred in (i)), and (iii) an SEM photograph of the same particle.

- A copper particle was passed through the transducer at different speeds over the possible mechanical speed range (0.4m/s to 4.0m/s) and the resulting pulse heights were found, to a very good approximation, to be independent of speed.
- The experimental V_{FMo}-V_{AMo} loci for particles of the seven materials mentioned above have been presented.

CHAPTER SEVEN: REFERENCES

1. Hansson C B: 'Physical data book in SI units', Permagon Press, Oxford (1972).

2. Woolman J and Mottram R A: 'The mechanical and physical properties of the British Standard steels', Macmillan, NY USA (1964).

3. ---: 'The electronics book 1981-1982', (manufacturers data book), ITT Electronic Services, Harlow, UK.

4. Standley K J: 'Oxide magnetic materials', Clarendon Press, Oxford, 2nd ed (1972).

CHAPTER EIGHT

DISCUSSION

In this chapter we compare results obtained from calibrating and testing the system, with the theory derived previously. Also, we relate the characteristics of the system to those required of an ideal wear-debris monitoring system and discuss the limitations of the present system.

8.1 A COMPARISON BETWEEN THE THEORY AND THE EXPERIMENTAL RESULTS

In this section, we compare the theory and the experimental results obtained, with reference to pulse shapes, the overall system constants and the system output locus diagram.

8.1.1 Pulse Shapes

In Chapter 3 we illustrated the theoretical pulse shapes generated by a particle passing through the inductive coil transducer. In Chapter 7 we showed plots of actual system output pulses due to the same situation. It is worth noting the similarity between the theoretical pulse shape for the 6mm long, 6mm diameter coil used in the system, and the pulse shapes obtained experimentally.

A pulse can be described in terms of various parameters, the most important of these being height and duration. The duration of the pulse, for a given particle, is proportional to the speed of the particle through the coil whilst pulse height should be constant over the designed speed range. In our system, operating within the approximate speed range range 0.4m/s to 4.0m/s produced almost constant pulse heights, as described in §7.3.1.

8.1.2 Overall System Constants

In §6.3 we obtained estimates for the system constants ($K_{FM} \simeq -1.47 \times 10^{11}$ V/m^3 and $K_{AM} \simeq 26.3 \times 10^{11}$ V/m^3) which, for a particle of known characteristics, yield the corresponding estimates for V_{FMo} and V_{AMo}, the system output pulse heights. However, we found that by calibrating with lead particles, the values of the system constants obtained ($K_{FM} \simeq -1.18 \times 10^{11}$ V/m^3 and $K_{AM} \simeq 5.9 \times 10^{11}$ V/m^3) were somewhat different to the estimates. The practical values of K_{FM} and K_{AM} were clearly lower than expected and the differences were probably due to:

(i) The value of K_{co}, the coil constant, being less than the theoretical estimate because a significant proportion of the the inductance of the oscillator tuned circuit was not due to the transducer. This fact alone could explain the low K_{FM} value.

(ii) The values of Q^* and $(G^*/V_o.dV_o/dG^*)$ which were used to evaluate K_{AM} probably being higher than in reality. The values of Q^* and G^* used did not take into account any additional loss mechanisms, such as the resistance of the coaxial cable which connects the transducer to the remainder of the marginal oscillator circuitry or induced eddy currents in the screening box which surrounds the transducer.

(iii) The estimated value of $(G^*/V_o.dV_o/dG^*)$ being obtained using an HPSPICE computer simulation programme which required, as input data, transistor parameters which could only be roughly estimated for operation at 10MHz.

The above sources of error were considered too difficult to quantify and, for subsequent system tests, we used the values K_{FM} and K_{AM} obtained from calibrating with lead particles.

8.1.3 The System Output Locus Diagram

We have plotted, in Figure 8.1.3(a), theoretical locus plots for V_{FMo} against V_{AMo} for various materials over the diameter size range 100-400μm using system constant values $K_{FM}=-1.18\times10^{11}$ V/m^3 and $K_{AM}=5.9\times10^{11}$ V/m^3. On the same diagram, we have plotted experimental points obtained from passing particles through the system transducer.

For the three non-ferrous materials tested (copper, lead and brass), we obtained good fits to the corresponding theoretical curves although it should be noted that the values of K_{FM} and K_{AM} used were obtained using the lead particles. The theoretical curves correspond to spherical particles whilst the experimental points are due to randomly shaped pieces and it seems reasonable to attribute most of the scatter around the theoretical curves to the random shapes. In addition, some scatter is probably due to particles being off the coil axis.

Results for three ferrous materials, ferrite, EN5A steel and EN9 steel, have also been plotted on Figure 8.1.3(a). (The points corresponding to mu-metal particles were not included because a theoretical curve could not be drawn due to the unavailability of data on the values of μ_r and σ for the material.) The theoretical curve for ferrite follows a straight line along the V_{FMo}-axis and the corresponding experimental points

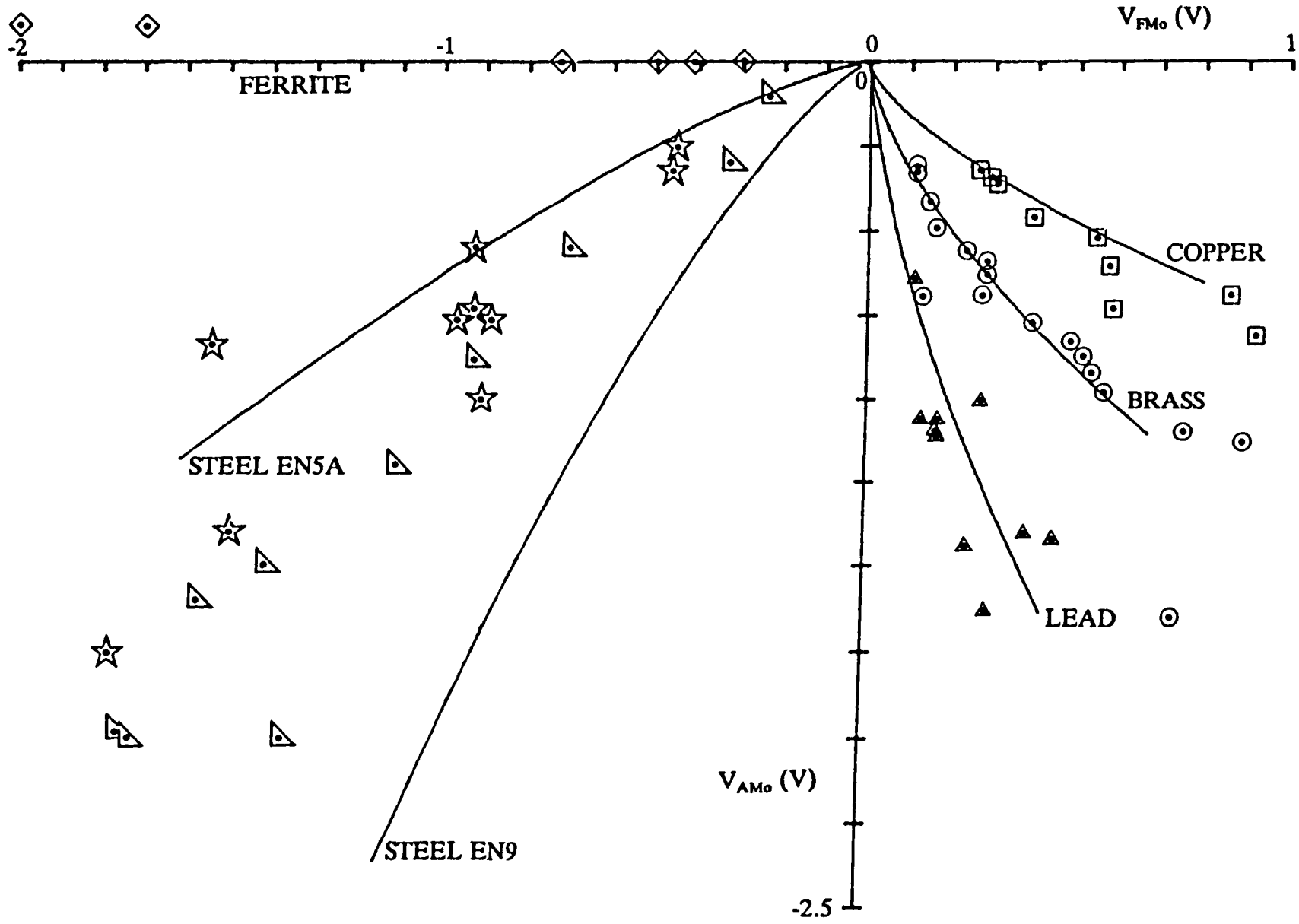


Figure 8.1.3(a) The system ouput locus diagram. Experimental results from six materials have been plotted and corresponding theoretical curves (assuming overall system constant values $K_{FM}=-1.18\times10^{11}$ V/m^3 and $K_{AM}=5.9\times10^{11}$ V/m^3) for the size range 100μm to 400μm, have been added. The experimental points are labelled by the symbols:

ferrite- ◇ copper- ⊡
steel EN5A- ☆ brass- ⊙
steel EN9- ◺ lead- ▲

lie quite close. Also, the EN5A steel experimental points fit their theoretical curve reasonably well. However, the EN9 steel experimental points produced a poor fit to the corresponding theoretical curve but this could be due to the value of μ_r used being too low. Had a value of 400 been used instead of 100, the curve would have fitted the results much better.

8.2 THE POTENTIAL FOR APPLICATION OF THE SYSTEM TO WEAR-DEBRIS CONDITION MONITORING

In this section, we compare the characteristics of the system developed to those required for wear-debris condition monitoring. In addition we compare the performance of the system with current wear-debris monitoring systems, both commercially available and under development.

8.2.1 Characteristics of the System in Comparison to the Characteristics Required of an Ideal Wear-Debris Condition Monitoring System

In §2.5.1 we outlined the features of an ideal lubricant wear-debris monitoring system and it is worthwhile to compare the features to the characteristics of our system:

(i) The system operates on-line and uses an electromagnetic transducer as suggested.

(ii) Manufacture of the system would, in principle, be quite straightforward since the inductive coil transducer is the only non-electronic component. Maintenance of the system would, for the same reason, be relatively simple.

(iii) Installation of the system would normally be easy, the only potentially difficult task being the insertion of the inductive coil transducer into the lubricant flow pipe-work.

(iv) The system allows us to detect and identify the material of non-ferrous metal particles in the approximate size range 200μm to 500μm whilst, for ferrous metal particles, the approximate range is 100μm to 400μm. This falls considerably short of the ideal requirement for detection over the size range 1μm to 1mm but, nevertheless, we can deal with particles in size ranges which, according to some previous workers, correspond to the onset of serious wear processes.

(v) The system is able to detect and identify the material of the particles while an indication of particle concentration could be obtained by counting over a period. Particle shape is not generally identifiable although it *may* be possible to infer a degree of shape variation from the extent of scatter of experimental points around

the theoretical V_{FMo}-V_{AMo} curve for a given material.

Thus, the original specifications and operational requirements for the wear debris monitoring system have, to a large extent, been achieved.

8.2.2 Characteristics of the System in Comparison with the Characteristics of Commercially Available On-line Wear-Debris Monitoring Systems

In §2.2.1 we described various on-line wear-debris monitoring techniques which are commercially available and it is worth comparing the characteristics of the system to these.

The Smiths Industries particle detector operates on-line and uses inductive coil transducers which form the arms of an AC impedance bridge. As explained in §4.1.1, bridge circuits only allow discrimination between ferrous and non-ferrous particles and cannot indicate, simultaneously, an estimate of particle size and material as can be done with marginal oscillator techniques.

The system developed here operates with a higher degree of sensitivity to small particles than the Smiths Industries product. However, the transducer in our system is considerably smaller in diameter (6mm compared with 12mm for the Smiths system) and we would naturally expect higher sensitivity.

Other commercial techniques already described include those dependent on optical principles, oil opacity being a major drawback, and on special filters, which may be unreliable and require frequent maintenance. The system developed in this project does not exhibit such problems.

8.2.3 Characteristics of the System in Comparison to the Characteristics of On-line Wear-Debris Monitoring Techniques Currently Being Researched

As described in §2.3, two systems currently being researched are an improved version of the Smiths Industries detector mentioned above and an ultrasonic technique.

The Smiths Industries detector is being improved through an uprated electronics package and is claimed to be capable of detecting 100μm iron particles. This corresponds to a level of sensitivity similar to that of our system.

The ultrasonic technique, due to Harries et al, indicates a sensitivity down to 35μm particles although we suspect that vibration in real machinery would result in significant degradation of performance. The technique seems unable to identify specific materials and depends on the particle being relatively hard.

8.3 THE LIMITATIONS OF SYSTEM PERFORMANCE

The performance of the system is limited by various effects; some due to necessary design compromises and others due to physical problems such a vibration and noise.

The system has been designed to detect particles moving through the transducer in the approximate speed range 0.4m/s to 4.0m/s with low-pass and high-pass filtering being applied to remove drift and noise effects whilst minimising distortion of signal pulses. The speed range may be regarded as a system limitation.

The heights of the output pulses, V_{FMo} and V_{AMo}, are related by the system proportionality constants K_{FM} and K_{AM}, to values of Re[D] and Im[D] for the particle passing through the transducer provided that the particle size is, for non-ferrous material, less than ~500μm or, for ferrous material, less than ~400μm in diameter. Larger particles are detectable, but nonlinearities in the various system components makes estimates of particle size and material become increasingly inaccurate.

For particles close to the lower size limits, the effect of noise makes identifying and measuring signal pulses difficult; the SNR (defined previously as the ratio of the system output pulse height to the rms noise voltage signal) rapidly worsens as the particle size is reduced. The minimum height which we felt was worthy of analysis was about twice the peak-to-peak noise voltage which, in turn, was about six times the rms noise voltage. Thus, an SNR of less than 12 (i.e. about 22dB), which corresponds to a ferrous particle of ~100μm or a non-ferrous particle of ~200μm, was considered too low for analysis.

As explained in §7.2.2, the system noise voltage was due to two main effects: electronic noise and microphonic effects. Electronic noise at the system outputs is probably due mainly to noise effects within the marginal oscillator which, in turn, are principally dependent on the noise characteristics of the active components of the feedback amplifier.

The other main source of system noise is due to microphonic effects which result from vibrations produced by the mechanical reciprocating mechanism used to test the system (see §7.1.5). The V_{FM} output signal is affected more than the V_{AM} output and we believe that modulation of the marginal oscillator tuned circuit capacitance is the principle cause. This is due to a combination of vibration in the transducer coil windings and in the other circuit components, especially the coaxial cable which connects the transducer to the remainder of the oscillator circuitry.

8.4 SUMMARY AND CONCLUSIONS FOR CHAPTER EIGHT

This chapter is summarised and concluded as follows:

- Pulse shapes generated at the system outputs in response to a particle being passed through the transducer are similar to the shapes obtained theoretically.
- The values for the overall system constants, K_{FM} and K_{AM}, obtained from calibrating using lead particles were lower than expected. Possible reasons for this have been discussed.
- Theoretical and experimental results for six materials have been plotted on a V_{FMo}-V_{AMo} locus diagram. In general, the theoretical and experimental points correlate well.
- The system exhibits most of the features desirable in an ideal on-line wear-debris monitoring system.
- The system potentially offers significant improvements over commercially available on-line wear-debris monitoring systems.
- The system exhibits advantages over other on-line wear-debris monitoring systems currently being researched.
- The present system cannot be used to analyse particles less than ~100μm in diameter for ferrous materials or ~200μm for non-ferrous materials because of noise. Particles greater than ~400μm in diameter for ferrous materials or ~500μm for non-ferrous materials cannot be analysed because of nonlinear distortion in the various system components.
- A major proportion of the system noise is due to vibration being transferred from the mechanical reciprocating mechanism to the transducer coil windings and other circuit components, particularly the coaxial cable which connects the transducer to the remainder of the system.

CHAPTER NINE

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER WORK

In this final chapter we state the conclusions for the project and recommend possible directions of further work.

9.1 CONCLUSIONS

- The use of condition monitoring techniques has been identified as a valuable aid to the maintenance of machinery.
- Various techniques for condition monitoring exist and include wear-debris monitoring which involves analysing the lubricant of a machine to detect and identify wear-debris particles and hence determine the 'health' of the machine.
- The nature of wear-debris has been reviewed. In general, the higher the concentration and size of wear-debris particles, the more severe the wear process. Wear debris monitoring techniques, either commercially available or at the research stage, are able to detect particles over various particle size ranges but suffer from an assortment of limitations. Moreover, particle analysis techniques from other fields are not easily applicable to wear-debris monitoring.
- The characteristics of an ideal wear-debris monitoring system have been identified but are not, in general, feasible for practical implementation. A realistic set of project objectives was therefore obtained and a prototype wear-debris monitoring system was designed, built and successfully tested.
- The use of an electromagnetic transducer was considered to be the best approach and an inductive coil transducer selected as the most suitable type. The theory of the interaction between metal particles and such a transducer has been derived, the coil series resistance and inductance being altered due to a metal particle interacting with the coil magnetic field. The nature of the interaction is strongly dependent on the frequency of operation and particle size and material. To obtain maximum information on a particle, the frequency of operation must ensure that induced eddy currents within the particle are partially skin-depth limited. For this mode of operation, computer simulations generating a locus plot of Re[D] versus Im[D], with particle size as the parametric variable, show that different materials follow separate loci. (The complex D value is a function of particle size, resistivity and magnetic permeability as well as the operating frequency.) Operation at ~10MHz was found to produce well-spaced loci for materials of

interest and this frequency was chosen for this work.

- Provided that the particle is small compared to the transducer geometry, the particle can, for the purpose of analysis, be considered a magnetic dipole. Theoretical values of Re[D] and Im[D] for a particle can be obtained by measurement of the change in coil impedance using simple relations which link a change in coil inductance to the value of Im[D] and a change in coil resistance to the value of Im[D]. In addition, ensuring that the particle is excluded from any region close to the coil windings whilst passing it through the coil produces pulses in coil inductance and resistance values.
- The most suitable method for obtaining the values of the coil inductance and resistance changes was considered to be the marginal oscillator method in preference to bridge or decay methods; the commonly-used twin-JFET configuration has been used in this work. The coil transducer forms the inductance in the oscillator tuned circuit and, to a good approximation, perturbations in the coil resistance amplitude modulate the oscillator signal whilst perturbations in coil inductance result in frequency modulation.
- The magnitude of the pulse changes in oscillator frequency and amplitude generated by a particle passing through the coil can be obtained by frequency and amplitude demodulation, and subsequent scaling by system constants K_{FM} and K_{AM} yields estimates of the values of Re[D] and Im[D] for the particle. An amplitude demodulator circuit based on an envelope detector was used whilst for FM demodulation, a circuit based on a monostable pulse demodulator was selected.
- Theoretical values of K_{FM} and K_{AM} were derived by considering the transducer geometry and circuit component values. However, calibrating the system using lead particles yielded a slightly different value for K_{FM} and a substantially different value for K_{AM}. Reasons for the differences have been discussed.
- Tests on the system included passing a copper particle through the coil transducer at varying speeds between 0.4m/s and 4m/s and this yielded pulses of constant height whilst the pulse width was found to be inversely proportional to the speed of the particle, as expected.
- Using the values of K_{FM} and K_{AM} obtained from calibrating the system with lead particles, we plotted theoretical and experimental points for six materials: lead, brass, copper, steel EN9, steel EN5A and ferrite. The theoretical and experimental results were generally in good agreement. Moreover, by plotting the

experimental values of Re[D] and Im[D] of a particle on the theoretical D loci plane, estimates of the particle size and material could be obtained.

- Noise levels in the system prevent analysis of particles of non-ferrous materials smaller than ~200μm, whilst for ferrous materials the lower limit is ~100μm. Distortion within the system limits the analysis to a maximum size for non-ferrous particles of ~500μm, whilst ferrous particles are limited to ~400μm.
- The project was aimed specifically at wear-debris monitoring and could, in its present form, be applied for that purpose. The limitations of the present system (see §8.3) would prevent detection and analysis of relatively small particles (e.g. steel less than ~100μm diameter) which could indicate the onset of severe wear. Nonetheless, the present system could be applied to large machines where relatively large particles may appear in the lubricant flow, before severe wear begins. In a wider sense, the present system could be applied in any circumstance where a non-optical non-contact technique for identification of metal particles is required, provided that they are suspended in a non-conducting, non-magnetic medium.

9.2 RECOMMENDATIONS FOR FURTHER WORK

Further work could commence with the application of the present prototype system to an operational machine but, initially at least, it would probably be better to set up a lubricant flow-rig within a laboratory environment where further tests could be made on the system. Moreover, improvements to the present system could be validated more conveniently and accurately using the flow-rig approach, as well as by making further use of the present mechanical test-rig.

A lubricant flow-rig would have to be capable of pumping the lubricant around a closed loop into which wear particles, either real or simulated, could be injected. The rig would have to be designed to allow attachment of the system transducer and be easy to clean.

Re-design of the present system could provide improved characteristics; the same basic configuration (i.e. inductive coil transducer → marginal oscillator → FM and AM demodulators) should still, however, be retained. Spurious noise due to microphonic effects would be reduced by, for example, stronger bonding of the coil windings of the transducer. Also, the marginal oscillator circuit would be improved by selecting lower noise active components. On the other hand, the demodulator circuits should probably not be changed as their performance is, at present, adequate.

The system output pulse signals can be digitised using analogue to digital converter circuits. The digitised signals can then be fed into a computer to analyse the pulses and produce more direct information on the size, material and concentration of particles. Preliminary work on such a scheme is currently being undertaken in the Department and further work on this area is recommended.

APPENDIX ONE: THE MAGNETIC FLUX DENSITY INSIDE THE COIL AT PLANE z=0 FOR VARYING RADIUS r

First, we quote an expression for the magnetic vector potential† at a point P at (r,0,0) within an (r,ϕ,z) cylindrical coordinate system due to an elemental ring as shown in Figure A1(a):

$$dA_\phi = \frac{\mu_o n I_c A dz}{4\pi} \int_{\phi=0}^{2\pi} \frac{\cos\phi}{[A^2 + r^2 + z^2 - 2A r\cos\phi]^{1/2}}\, d\phi \qquad (1)$$

The value of the total vector potential due to the coil is obtained by integrating, over the coil length, the potentials due to all such elementary rings to produce

$$A_\phi = \frac{\mu_o n I_c A}{4\pi} \int_{z=-\frac{\lambda}{2}}^{\frac{\lambda}{2}} \int_{\phi=0}^{2\pi} \frac{\cos\phi}{[A^2 + r^2 + z^2 - 2A r\cos\phi]^{1/2}}\, d\phi\, dz \qquad (2)$$

and this expression can be evaluated numerically by repeated application of a numerical intergration algorithm.

Using the relation $\mathbf{B}=\nabla\times\mathbf{A}$ within our cylindrical coordinate system yields

$$\mathbf{B} = B_z \mathbf{a}_z = \left[\frac{\partial A_\phi}{\partial r} + \frac{A_\phi}{r}\right] \mathbf{a}_z \qquad (3)$$

and this may be obtained numerically from the result of Equation (2).

† Solymar L: 'Lectures on electromagnetic theory', 2nd ed, Oxford University Press, Oxford (1984), § 3.7.

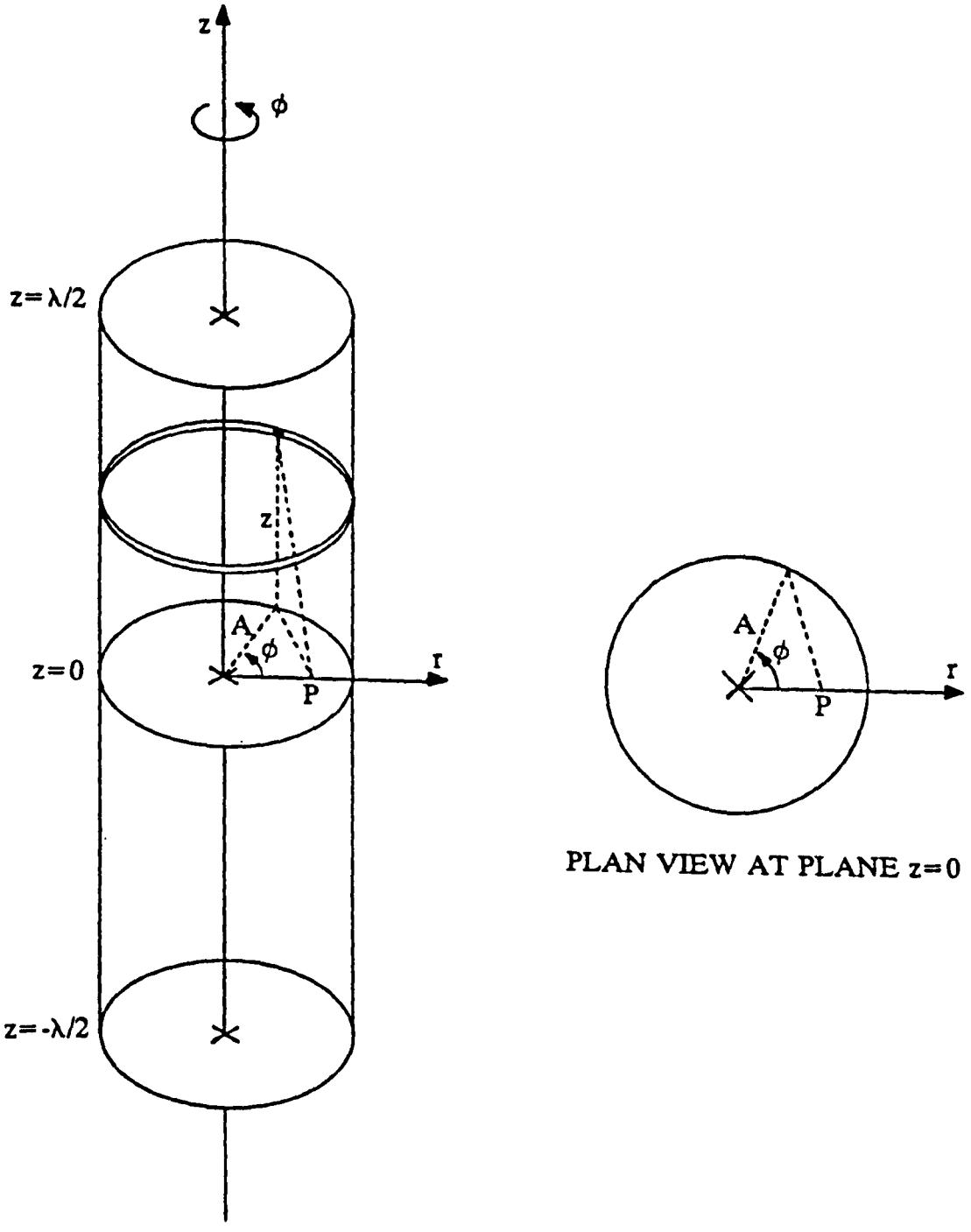


Figure A1(a) Geometric model of coil represented by a current sheet, for calculation of magnetic flux density at point P.

APPENDIX TWO: THE FLUX-LINKAGE GENERATED BY A MAGNETIC DIPOLE AT PLANE z=0 FOR VARYING RADIUS r

First, we quote an expression for the magnetic vector potential field due to a magnetic dipole of strength m at a distance much greater than the dipole radius, within an (R,θ,z) cylindrical coordinate system†:

$$A_\theta = \frac{\mu_o m}{4\pi} \frac{R}{[R^2 + z^2]^{3/2}} a_\theta \tag{1}$$

Consider such a dipole positioned at point P in Figure A2(a). The flux linking an elemental coil loop is given by the line integral

$$\Phi(z) = \oint A_\theta . dl \tag{2}$$

which may be evaluated around a single loop:

$$\Phi(z) = \oint A_\theta . dl$$

$$= \int_{\phi=0}^{2\pi} \frac{\mu_o m}{4\pi} \frac{R}{[R^2 + z^2]^{3/2}} \cos(\theta - \phi)\, A\, d\phi \tag{3}$$

where, from Figure A2(a),

$$R = [(A\sin\phi)^2 + (A\cos\phi - r)^2]^{1/2} \tag{4}$$

and

$$\cos(\theta - \phi) = \frac{A^2 + R^2 - r^2}{2Ar} \tag{5}$$

Hence, by substituting Equations (4) and (5) into (3)

$$\Phi(z) = \int_{\phi=0}^{2\pi} \frac{\mu_o m}{4\pi} \frac{(A\sin\phi)^2 + (A\cos\phi - r)^2 + A^2 - r^2}{2[(A\sin\phi)^2 + (A\cos\phi - r)^2 + z^2]^{3/2}}\, d\phi \tag{6}$$

The flux linkage due to the dipole interacting with the entire coil is given by

$$\Lambda_p(r,0) = \int_{z=-\lambda/2}^{\lambda/2} \Phi(z)\, n\, dz$$

$$= \int_{z=-\lambda/2}^{\lambda/2} \int_{\phi=0}^{2\pi} \frac{\mu_o mn}{4\pi} \frac{(A\sin\phi)^2 + (A\cos\phi - r)^2 + A^2 - r^2}{2[A\,(A\sin\phi)^2 + (A\cos\phi - r)^2 + z^2]^{3/2}}\, d\phi\, dz \tag{7}$$

† Solymar L: 'Lectures on electromagnetic theory', 2nd ed, Oxford University Press, Oxford (1984), §3.7.

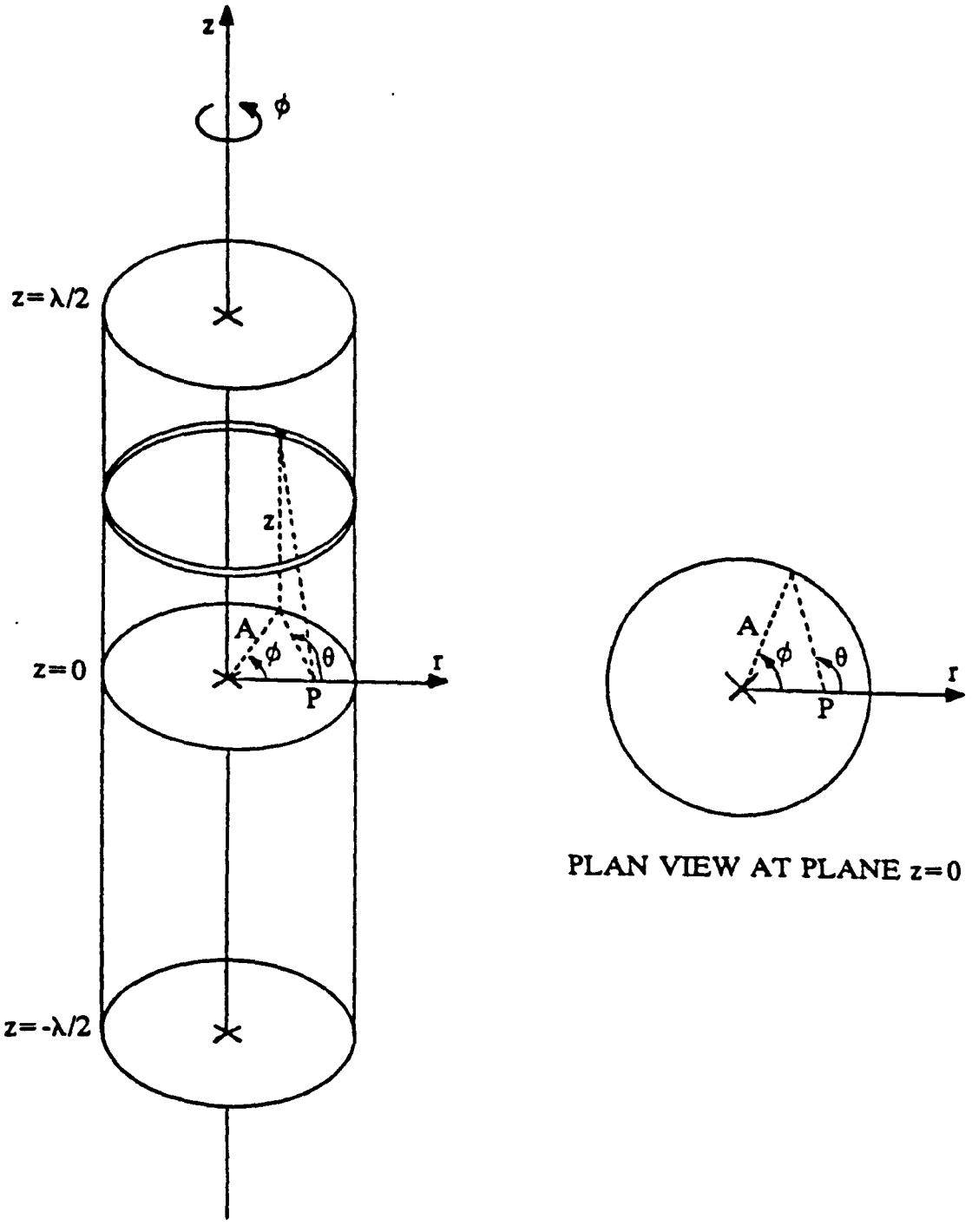


Figure A2(a) Geometric model of coil represented as a current sheet for calculation of flux-linkage due to a magnetic dipole positioned at point P.